Il peccato originale delle intelligenze artificiali

L'origine degli stereotipi di genere
attraverso "fogli illustrativi"

Il peccato originale delle intelligenze artificiali
L'origine degli stereotipi di genere attraverso "fogli illustrativi"

LAUREANDA Beatrice Bazzan

RELATRICE Valeria Bucchetti

CORRELATRICE Francesca Casnati

FONT GT Sectra, by Dominik Huber, Marc Kappeler (Moiré), Noël Leu (Grilli Type).
PP Neue Montreal, by Mat Desjardins (Pangram Pangram).

Laurea magistrale in Design della Comunicazione
Scuola del Design, Politecnico di Milano
A.A. 2019/2020

Index

Abstract

ITA

Gli algoritmi sono ai più incomprensibili e in tal senso trattati come universalmente veritieri. Per esistere si nutrono di dati che sono però scelti, selezionati e filtrati da esseri umani e infine da questi ultimi implementati. Nel processo si annidano molteplici bias, tra cui spiccano per numero i *gender bias*. La tesi vuole indagare come la dimensione del progetto di comunicazione possa supportare la ricerca e lo sviluppo di intelligenze artificiali più eque ed inclusive. Avvalendosi della lente di decostruzione del design e degli studi di genere, il fine ultimo è fornire uno strumento che possa individuare la presenza di gender bias nei contenuti visivi di allenamento delle intelligenze artificiali.

ENG

Algorithms are to most people incomprehensible and in that respect treated as universally true. In order to exist, they feed on data that are, nevertheless, chosen, selected and filtered by humans, and implemented by them. In this process multiple biases lurk, of which gender bias is the most frequent. The dissertation aims to investigate how the dimension of the communication design can support the research and development of fairer and more inclusive artificial intelligences. Using the deconstruction lens of design and gender studies, the ultimate goal is to provide a tool that can analyze and detect the presence of gender bias in the visual content of IA training.

Introduzione

La storia della *computer vision*, settore dell'intelligenza artificiale che insegna alle macchine come interpretare le immagini, inizia con una leggenda metropolitana. Erano giorni di calura estiva del 1966, mentre ci si avvicinava alla pausa estiva al MIT, quando un giovane professore di nome Marvin Minsky decise che l'abilità di interpretare le immagine fosse una funzione principale per l'IA. Per questo Minsky si rivolse prontamente a Gerald Sussman, suo studente, fornendogli un compito: avrebbe dovuto passare l'estate a collegare una telecamera ad un computer, facendo in modo che il computer descrivesse cosa stesse vedendo. Questo diventò poi il famoso "Summer Vision Project"[1]. Ovviamente progettare un sistema che consentisse ai computer di "vedere" non era un compito semplice e non durò una sola estate del 1966, ma si protrasse fino agli anni 90, quando la scoperta del modello probabilistico e delle tecniche di insegnamento accelerarono il processo che continua ancora tutt'oggi, sebbene varie sfide quali il rilevamento di oggetti e il riconoscimento facciale siano stati quasi del tutto risolti.

L'inevitabilità ricorre in molte narrazioni dell'IA: si presume che i miglioramenti tecnici risolveranno tutti i problemi e le limitazioni. Viene da chiedersi, però, se la sfida di interrogare le macchine per descrivere "cosa vedono" potrà mai essere effettivamente risolta.

La tesi osserverà come l'interpretazione automatica delle immagini sia un progetto complesso e non puramente tecnico: quest'ultima deve infatti coinvolgere una parte umanistica imprescindibile, che non appartiene alle IA.

I sistemi di IA sono ormai ubiqui nell'advertising, nelle assunzioni, nei servizi di finanza, nelle policing ed in molti altri campi, e possono perpetuare disuguaglianze sociali esistenti, anche basate sul genere. Per questo risulta più importante che mai, data l'espansione costante, capire quali meccanismi siano alla base e capire come il designer della comunicazione possa intervenire nel processo. Comprendere quindi il motivo per il quale molti computer collegano automaticamente la donna al ruolo di "casalinga" e l'uomo a quello di "professionista".

1 Il professore Minsky ottenne il credito per l'idea, ma Sussman, insieme al resto del team che svolgeva i compiti estivi è stato fondamentale per questi primi sviluppi della computer vision. Vedi Seymour A. Papert, "The Summer Vision Project", 1 luglio 1966, al link: <https://dspace.mit.edu/handle/1721.1/6125>. Come scrisse: «The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system».

La tesi vuole risalire al “peccato originale delle IA”: comprendere il problema che porta all’iniquità ed alla disuguaglianza nei risultati ottenuti dai sistemi algoritmici. Per farlo è necessario rivolgersi al “nutrimento” dei sistemi, ossia ai dataset di allenamento:

> C’è molto in gioco nell’architettura e nei contenuti degli insiemi di allenamento usati nell’IA. Possono promuovere o discriminare, approvare o rifiutare, rendere visibile o invisibile, giudicare o imporre. E così abbiamo bisogno di esaminarli - perché sono già usati per esaminare noi - e di avere una discussione pubblica più ampia sulle loro conseguenze, piuttosto che tenerla all’interno dei corridoi accademici. Poiché i set di formazione sono sempre più parte delle nostre infrastrutture urbane, legali, logistiche e commerciali, hanno un ruolo importante ma sottovalutato: il potere di modellare il mondo nelle loro stesse immagini. (Crawford e Paglen 2019)

Il fine di questa tesi è l’individuazione delle *best practices* che un designer può espletare in relazione ad un progetto di *machine learning*, dove ogni singola informazione data come *input* restituisce un *output*. Se l’input contiene bias, l’output non solo li ripeterà ma li amplificherà: un famoso detto tra gli informatici è infatti «garbage in, garbage out[2]».

L’obiettivo ultimo è la progettazione di un “foglio illustrativo” per verificare la presenza di gender bias nei dataset di allenamento, applicando – cosa non consueta – il punto di vista del designer della comunicazione. Verranno in questo senso analizzate le immagini sia in modo qualitativo che quantitativo, per verificare la ricorrenza di stereotipi di genere nella rappresentazione di ruoli e azioni.

Il foglio illustrativo deve essere semplice, aggiornabile ed utile a coloro che:

- usano i dataset, ovvero i programmatori e *data scientist* che implementano i sistemi di ML;
- creano i dataset;
- ricercano sul tema dei *bias* nell’IA;
- deve inoltre essere spunto di riflessione per il pubblico più vasto.

2 *Garbage in, garbage out* (GIGO), nel contesto della tecnologia dell’informazione, è un’espressione gergale che significa indipendentemente dall’accuratezza della logica di un programma, i risultati saranno errati se l’input non è valido. Mentre il termine viene usato più frequentemente nel contesto dello sviluppo del software, GIGO può anche essere usato per riferirsi a qualsiasi sistema decisionale in cui la mancata presa di decisioni giuste con dati precisi e accurati potrebbe portare a risultati errati e insensati.

Questi quattro identificano il target dell'*output* progettuale che mira ad offrire una metodologia di ricerca, insieme ad una visione di come gli stereotipi di genere si inseriscono e vengono, solitamente in modo involontario, codificati all'interno delle macchine.

L'idea nasce da un *corpus* di studi in ambito di scienza dei dati che si propone di utilizzare delle metodologie uniche e comuni per la progettazione e l'analisi degli algoritmi. L'obiettivo è costruire una sorta di etichetta degli ingredienti costitutivi dei dataset, organizzata in modo rigoroso. L'etichetta si propone di diventare una documentazione da affiancare al dataset, utile a capirne gli aspetti fondamentali e ad implementarne la trasparenza.

L'approccio che la tesi si propone di compiere è una *summa rerum*, una descrizione primaria del dataset basato su domande di ricerca che provengono dal corpus sopra citato, cui segue un modulo di ampliamento focalizzato sui contenuti visivi che vengono dati in pasto alle IA e da cui queste ultime "imparano".

La struttura dell'elaborato si articola in sei capitoli.

» Il capitolo 1 fornisce un quadro generale delle IA, della loro importanza, dello sviluppo e della sempre maggiore implementazione. Da qui origina il percorso che porta alla comprensione dei bias, sia negli umani che nei sistemi algoritmici e di come, in particolare, questi si insinuino nei dataset e successivamente nei risultati delle IA. Il percorso si sofferma su una tappa fondamentale: sfata il mito della tecnologia come imparziale. Il capitolo si conclude con un appunto e un monito: i bias non sono così facili da estirpare, c'è quindi bisogno di progettare rimedi all'infestazione.

» Il capitolo 2 introduce l'importanza del design nella problematica analizzata. La ricerca si avvale di una lente di analisi basata sul femminismo intersezionale, punto di partenza per uno studio femminista dei dati, alla ricerca della co-liberazione. Il capitolo cerca di spiegare

l'origine delle diseguaglianze di genere, prima nella società e a seguire nelle IA. Le cause a cui la ricerca risale sono principalmente tre: i numeri di genere molto sbilanciati del personale che si occupa dei sistemi di IA, la non rappresentazione dei dati e infine dati non esistenti sul genere non-standard. Innumerevoli sono le conseguenze e in questo capitolo ne vengono riportati solo alcuni esempi che cercano di perimetrare il problema.

» Il capitolo 3 inquadra il gender bias attraverso gli strumenti digitali, considerando gli algoritmi come un riferimento culturale per analizzare le disuguaglianze di genere. Il focus inquadra il raggio d'azione all'interno del quale il designer della comunicazione può ed è capace di intervenire: l'analisi delle immagini. Le figure sono da sempre veicolo di stereotipizzazione di genere e continuano ad esserlo anche nei sistemi di IA. Viviamo nella società dell'immagine, nella quale la loro diffusione influenza il pensiero, è veicolo di conoscenza ed ogni immagine incorpora e trasmette modi di vedere. Vengono qui presentati i principali metodi di analisi di immagini esistenti in letteratura.

» Il capitolo 4 rappresenta la svolta verticale ed introduce la soluzione che la tesi si propone di progettare sperimentalmente: una metodologia di analisi per i dataset di allenamento delle IA. Viene presentato qui il concetto di foglio illustrativo, riportando sei casi studio dei quali si illustrano le motivazioni, gli strumenti utilizzati, i formati. L'ultima parte del capitolo pone le basi per il successivo e individua i codici comunicativi attraverso i quali gli stereotipi di genere vengono rappresentati nelle immagini, quindi come organizzare l'analisi successiva.

» Il capitolo 5 entra nel merito del progetto, spiegando fase per fase in cosa consiste il foglio illustrativo, come vengono selezionate le materie prime di analisi (i dataset), i criteri di analisi, le metodologie e i tre livelli in cui il foglio il-

lustrativo si suddivide. La terza parte è quella su cui ci si sofferma maggiormente, essendo il vero e proprio apporto del designer nella decostruzione del sistema di senso delle singole immagini di cui i dataset analizzati si compongono. Il capitolo procede poi nella strutturazione e nei metodi di generazione del foglio illustrativo, insieme alla spiegazione delle scelte stilistiche e metodologiche di design. Prosegue nel verificare i risultati individuati, tra cui i principali *insight* che fuoriescono dall'analisi. Infine individua i limiti che hanno vincolato la ricerca e la progettazione.

1.0 Genesi: i bias nell'intelligenza artificiale

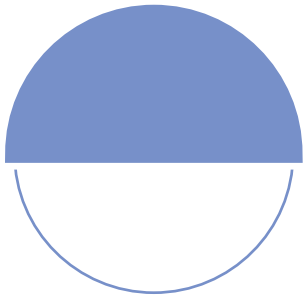
In many ways data is destiny. When you think of AI is forward looking. But AI is based on data. And data is a reflection of our history. So the past dwells within our algorithms.
Kantayya 2020

1.1 Perché è necessario far luce sui bias negli algoritmi

La nostra vita oggi è più virtuale che reale, tanto da dar senso all'espressione *onlife*. Data la velocità supersonica di questa evoluzione, è sempre più fondamentale riflettere su un «fenomeno che potrebbe paradossalmente condurre a una società onlife soprattutto maschile, nuovamente costruita "a misura di uomo", come è avvenuto nelle civiltà classiche o nelle rivoluzioni liberali» (D'Amico 2020).

È necessario approfondire i problemi insiti negli algoritmi perché questi ultimi sono oramai ubiqui. Fanno parte della nostra società non meno non meno di quanto ne fanno parte un amministratore delegato, una mamma, un papà ecc. Dalla ricerca di McKinsey (Balakrishnan et al. 2020) risulta che le aziende utilizzano l'IA per generare valore che, sempre di più, finisce per diventare il reddito degli stessi enti. Metà dei partecipanti al sondaggio ha risposto che la propria azienda ha adottato l'IA per almeno una funzione commerciale, tra cui spiccano le attività di servizio, lo sviluppo di prodotti, il marketing, le vendite e altri ambiti estremamente sensibili quali le risorse umane e il calcolo e la gestione dei rischi.

Parlando di *deep learning*, sono le aziende *high-tech* e di telecomunicazione a primeggiare nella corsa all'innovazione. Il 30% degli intervistati di questi settori affer-



50%

degli intervistati riferiscono che le loro aziende hanno adottato l'IA in almeno una funzione aziendale

Fonte: *The state of AI in 2020*, McKinsey Global Survey

FIG. 1 Il 50% delle aziende dichiara di aver implementato l'IA.

ma che le aziende hanno incorporato capacità di DL. Scendendo nei particolari del sondaggio, la percezione di quanto sia importante aggiornare i modelli di IA in base a criteri definiti è decisamente bassa: si riporta un tasso pari al 15% tra i *non high performance*[1] (da ora NHP), a confronto con un non molto più rincuorante 45% tra gli *high performance* (da ora HP). Le percentuali si alzano per quanto riguarda i protocolli per la qualità dei dati: si trova un 29% per gli NHP e un 48% per gli HP. Nonostante questi e successivi numeri riguardanti le *best practices* siano decisamente bassi, alla domanda che chiede quanto ci si senta a proprio agio nel far prendere decisioni alle IA relative agli investimenti risponde positivamente il 65% degli HP.

1 Il sondaggio di McKinsey definisce *high performance* le aziende per le quali almeno il 20% o più dell'EBIT nel 2019 era attribuibile al loro uso dell'IA.

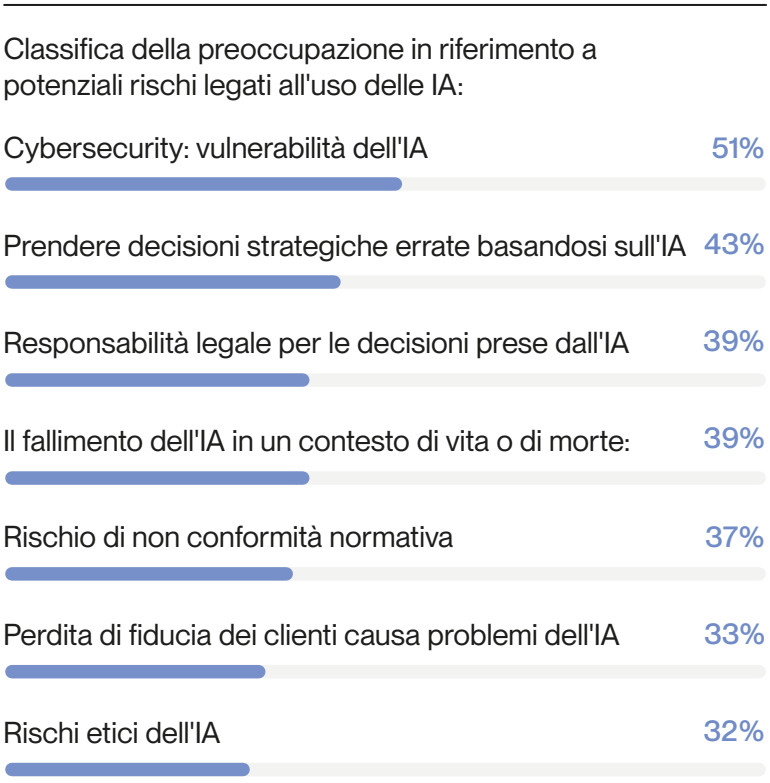
Il panorama non si differenzia molto osservando la parallela ricerca *Deloitte State of AI in the Enterprise, 2nd Edition,* 2018 di Deloitte Insights: è stato richiesto ai dirigenti intervistati di esprimere una percentuale di preoccupazione riguardo ai rischi che comporta l'IA. Il rischio "etico" è messo all'ultimo posto nella scala della preoccupazione, con uno scarso 32%.

I risultati dei sondaggi suggeriscono che solo una minoranza di aziende riconosce i rischi insiti nell'uso dell'IA e che un numero ancora più ridotto investe per ridurli. La *cybersecurity* è l'unico pericolo ritenuto rilevante per le organizzazioni degli intervistati.

Gartner prevede che entro il 2022, l'85% dei progetti di IA darà risultati errati a causa di distorsioni nei dati, degli algoritmi o presenti all'interno dei team responsabili della loro creazione e gestione (cfr. Teigland 2019).

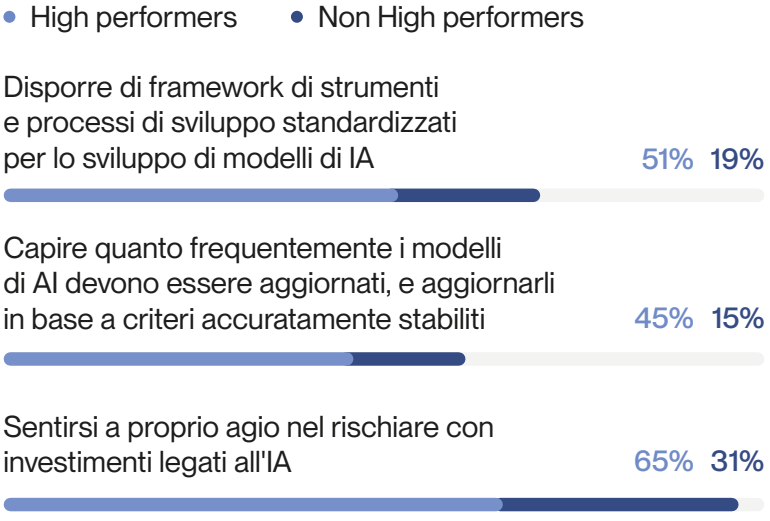
Il problema non riguarda solo la parte effettiva di progettazione del modello di IA ma coinvolge il dataset con cui il modello viene addestrato. I dataset sono un insieme di innumerevoli dati che vengono solitamente etichettati e classificati per singola immagine, singolo video o singolo file. Il processo è non di rado svolto da lavoratori a cottimo online (chiamati *Amazon turkers*[2]), pagati frazioni di centesimo. Questo lavoro porta spesso ad etichettature prive di senso, che cristallizzano inevitabilmente *unconscious bias*, i quali

2 Amazon Mechanical Turk: Access a global, on-demand, 24x7 workforce. Si tratta di un mercato di *crowdsourcing* per esternalizzare i processi e lavori aziendali a una forza lavoro distribuita, che può eseguire questi compiti virtualmente. Link: <www.mturk.com/>.



Fonte: *Deloitte State of AI in the Enterprise*, 2nd Edition, 2018.

FIG. 2 I principali rischi legati alle IA secondo il report *State of AI in the Enterprise, 2nd Edition* 2018 di Deloitte.



Fonte: *The state of AI in 2020*, McKinsey Global Survey

FIG. 3 Infografica che identifica le differenze tra aziende *high e non high performers,* individuate dal report *The State of AI in 2020* di McKinsey.

subentrano quando in un meccanismo di pensiero si vuole dare più valore alla rapidità che alla profondità.

Kate Crawford, parlando degli "etichettatori di dati" in *Anatomy of AI*, fa riferimento allo sfruttamento digitale. Il nome "turkers" nasce dall'invenzione dell'ungherese Wolfgang von Kempelen, che nel 1770 chiamò *Mechanical Turk* (Turco Meccanico) una macchina da lui costruita per giocare a scacchi (cfr. Crawford e Joler 2018), con l'obiettivo di impressionare l'imperatrice Maria Teresa d'Austria. Mechanical Turk ebbe uno strepitoso successo: vinse contro avversari umani per circa novant'anni. Si venne poi a scoprire che non si trattava di alcun sistema tecnologicamente avanzato: consisteva in realtà in una grande macchina, talmente spaziosa da permettere ad un maestro di scacchi umano di nascondersi all'interno. Centosessanta anni dopo Amazon chiama la sua piattaforma di *crowdsourcing* basata su micro-pagamenti riprendendo il nome di questa invenzione.

> Con Amazon Mechanical Turk, può sembrare agli utenti che un'applicazione stia usando l'IA avanzata per svolgere dei compiti. Ma è più vicino a una forma di "intelligenza artificiale artificiale" guidata da una forza lavoro clickworker remota e

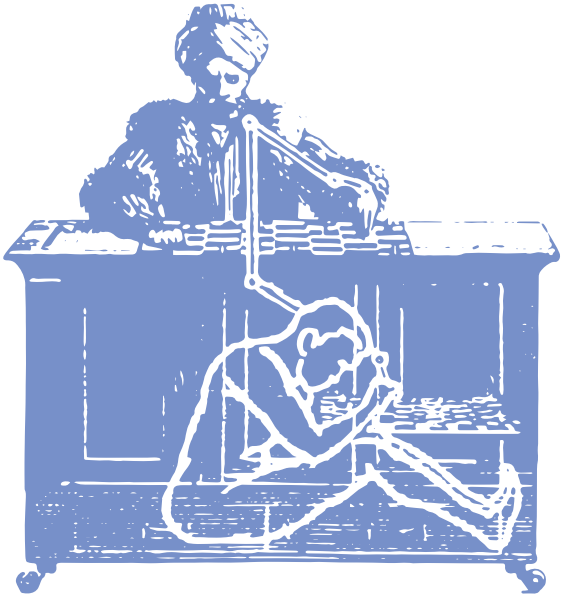


FIG. 4 Il turco meccanico.

> dispersa che aiuta un cliente a raggiungere i suoi obiettivi di business. Come osservato da Aytes, "in entrambi i casi [sia il Mechanical Turk del 1770 che la versione contemporanea del servizio di Amazon] la performance dei lavoratori che animano l'artificio è oscurata dallo spettacolo della macchina"[3]. (Crawford e Joler 2018)

3 Citazione di Aytes, *Return of the Crowds*.

Amazon Mechanical Turk è un lavoro che si fa carico dell'anonimato, che è invisibile all'interfaccia ed è nascosto dietro a processi algoritmici. Oggi la manodopera invisibile è un concetto comune: Google, attraverso *reCaptcha*, ne fa un lavoro non retribuito. Per poter dimostrare, infatti, di non essere un agente artificiale (si fa riferimento alla ormai nota spunta "I'm not a robot"), ogni utente della rete è costretto ad addestrare gratuitamente il sistema di riconoscimento delle immagini di Google selezionando le caselle che corrispondono alle immagini corrette.

Da questo punto di vista le forme odierne di IA si fanno un po' meno artificiali e molto più umane, considerato che la tecnologia è radicata e basata sullo sfruttamento di veri e propri corpi e cervelli umani.

FIG. 5 Calcolo e quantificazione della realtà naturale.

Tutto tende verso la missione infinita di mappare un'enciclopedia delle conoscenze umane: senza sosta si estraggono, si puliscono e si riordinano dati. La sete sfrenata di nuove risorse e di campi di sfruttamento cognitivo ha spinto verso la ricerca di strati sempre più profondi di ricerca della psiche umana, fino ad arrivare a mappare le emozioni attraverso l'espressione del viso. Esistono oggi dataset di training per rilevare le emozioni umane (ad esempio *AffectNet*, *Extended Cohn-Kanade Dataset*, *FER-2013*), per la somiglianza dei lineamenti (*Flickr Faces HQ Dataset Tufts Face Database*, *Real and Fake Face Detection*), per tracciare l'invecchiamento (*AgeGuess*), per le azioni umane (*AVA*, *OpenImages*, *MS COCO*). Ogni forma di dato biologico è in via di raccolta: dati forensi, biometrici, sociometrici e psicometrici.

Sembrerebbe di essere stati catapultati nel futuro, se non ci si fermasse ad osservare la qualità dei dataset. Questi ultimi si basano infatti su presupposti antichi e, come affermano Crawford e Joler, ripetono pattern sociali stereotipati e basati su dinamiche antiquate che si cerca di oltrepassare da tempo nell'*off-life*, nella vita fuori dalla rete. Un esempio: il dataset *AVA* mostra principalmente immagini taggate come donne alla voce "play with kids" e immagini taggate uomini in "kick (a person)" (cfr. Crawford e Joler 2018).

> The training sets for AI systems claim to be reaching into the fine-grained nature of everyday life, but they repeat the most stereotypical and restricted social patterns, re-inscribing a normative vision of the human past and projecting it into the human future. (*ibidem*)

Riflettere sulla composizione dei dataset risulta fondamentale poichè, come ha affermato Shalini Kantayya (attivista, regista e producer di "Coded Bias") alla Milano Digital Week[4], le IA oggi sono diventate dei veri e propri *gatekeepers* delle opportunità, in quanto decidono chi si deve sottomettere allo scrutinio della polizia, chi riceve un vaccino, chi riceve un prestito, una promozione, chi viene licenziato, ecc. Oramai le IA si fanno carico di decisioni che governano in gran parte il destino umano. Questi sistemi di cui ci fidiamo ciecamente non sono stati controllati da *sessismo* ne da *razzismo*. A volte non sono stati nemmeno controllati (prima

4 L'evento Milano Digital Week è avvenuto in data 20/03/2021. Link: <www.milanodigitalweek.com/tech-gender-bias>.

della loro implementazione) per certe forme di inaccuratezza, a meno che non si scontrino con il diretto guadagno delle aziende (cfr. Kantayya 2021).

1.2 Bias: di cosa si tratta

"Bias" è un termine tecnico, cui si ricorre in psicologia, statistica, informatica e svariate altre discipline. È difficilmente traducibile in italiano, dato che il significato dipende in modo indissolubile dal contesto di appartenenza. "Pregiudizio" è il suo più vicino corrispondente in italiano, ma tende a tralasciare un significato ben più ampio: «bias, infatti, oltre ad indicare comunemente un pregiudizio generalizzato, indica un errore sistematico» (Huyskes 2020). Il termine risulta quindi ampiamente adatto a descrivere gli stereotipi di genere nella loro totalità.

> Con "gender bias" si indica infatti un errore di classificazione (di matrice sociale) che porta a trattare le persone in modo iniquo e diverso in base al genere di appartenenza, e che colpisce soprattutto le donne (e i generi diversi da quello maschile dalla nascita). (*ibidem*)

Nell'informatica *algorithmic bias* descrive sia l'indicazione dei pregiudizi perpetrati dai sistemi algoritmici, sia la stima non corretta di dati o classifiche.

Una categoria di pregiudizi noti come pregiudizi cognitivi, sono modelli ripetuti di pensiero che possono portare a conclusioni imprecise o irragionevoli.

1.2.1 Tipi di bias cognitivi

Ci sono oltre 170 tipologie di bias cognitivi, suddivisibili in quattro classi (nonostante alcuni bias siano riconducibili a tutte e quattro). Lauren Isaacson, fondatrice di Curio Research, individua e definisce questi quattro contenitori in un suo *speech*[5].

Li definisce come:

5 Isaacson, L. (s.d.) dal video di Youtube *Boosting Your Bias Immunity*. Link: <www.xd.adobe.com/ideas/principles/emerging-technology/removing-ai-bias-pt-1-people-problem/>.

1 Informazioni necessarie (*Abundant information*). L'essere umano agisce seguendo una selezione filtrata di informazioni, dato che non è possibile conoscere ogni dettaglio di un argomento:
   - » euristica della disponibilità (*availability heuristics*): si tende a notare cose che già esistono nella mente grazie a esperienze passate;
   - » ancoraggio (*anchoring*): l'ordine con cui si ricevono le informazioni influenza il giudizio su di esse;
   - » bias di conferma (*confirmation bias*): si tende ad essere d'accordo con informazioni che corrispondono a quello a cui già si crede;
   - » effetto del bizzarro (*bizarreness effect*): si ricordano con più facilità eventi particolari.

2 Contesto limitato (*Limited context*). L'essere umano non può conoscere tutto per questo tende a riempire i buchi di conoscenza:
   - » effetto di inquadratura (*framing effect*): si giudicano gli oggetti e i soggetti per come essi sono presentati;
   - » effetto placebo (*placebo effect*): credere che qualcosa funzioni può essere tanto potente quanto qualcosa che funziona davvero;
   - » errore fondamentale di attribuzione (*fundamental attribution error*): si giudica l'altro per quello che si vede nel momento, giudichiamo noi stessi basandoci sulla situazione;
   - » pregiudizio nel gruppo (*in-group bias*): si tende a preferire le persone che appartengono al nostro stesso gruppo.

3 Tempo limitato (*limited time*): in situazioni in cui c'è necessità di agire in fretta si tende ad avere reazioni e pensieri immediati:
   - » bias dell'ottimismo (*optimism bias*): sovrastimare la possibilità di un risultato positivo;
   - » effetto Barnum (*Barnum effect*): connettere i concetti attraverso salti di immaginazione;
   - » *Dunning-Kruger*: quando non si è esperti di una materia si tende a pensare che la cono-

scenza comune corrisponda a tutto quello che c'è da sapere sull'argomento;
» pensiero di gruppo/effetto gregge (*group-think/bandwagon effect*): le opinioni si basano sull'essere d'accordo con il gruppo piuttosto che con le informazioni evidenti.

4 Memoria limitata (*Limited memory*). Non potendo ricordare tutto, solo certe informazioni tendono a consolidarsi nella mente:
» bias di negatività (*negativity bias*): gli eventi negativi e le emozioni che accompagnano questi restano con noi più a lungo rispetto agli eventi positivi;
» effetti di distanziamento (*spacing effects*): si impara meglio in piccole quantità distribuite nel tempo invece che tutto in una volta;
» stereotipo implicito (*implicit stereotype*): le associazioni imparate tra differenti qualità e categorie sociali;
» regola del climax (*peak-end rule*): le persone giudicano un'esperienza basandosi su come si sono sentiti nel suo punto più intenso, piuttosto che giudicarla nella sua interezza.

Le persone sono «naturalmente prevenute» (*Psychology Today*; traduzione mia). Dal principio della nostra vita assegnamo giudizi basati sulle prime impressioni. In sintesi: una persona non può essere imparziale, tutti abbiamo un certo grado di pregiudizio. Il pregiudizio cognitivo è un mezzo a cui si ricorre quando si ha bisogno di attivare un pensiero e una reazione rapida, a discapito dell'accuratezza che deriva invece dal pensiero profondo.

1.2.2 Tipi di bias negli algoritmi

Mentre i bias umani sono ben documentati da innumerevoli studi quali i test di associazione implicita (Greenwald, Mcghee e Schwartz 1998), gli esperimenti sul campo (Esperimento di Batson et al. 2002), le teorie psicologiche e sociali (Gruppi umani e categorie sociali, libro di

Henri Tajfel del 1981; Public Opinion, libro di Walter Lippmann del 1922), dei bias algoritmici si parla ma senza riferimenti univoci a teorie. Lo dimostra una ricerca su Google in cui, digitando la *query* "*type of bias in ai",* si trovano risultati completamente diversi: ogni sito, ogni testata ed ogni ricercatore ha un proprio metodo di elencare e distinguere i bias e sembra che non ci sia un accordo, né tanto meno una conoscenza di fondo comune convenzionata.

Nel 2017 Will Knight riporta in un articolo di Technology Review le parole delle due fondatrici della *AI Now Initiative*: Kate Crawford, ricercatrice di Microsoft, e Meredith Whittaker, ricercatrice di Google. In una *e-mail* scrivono: «It's still early days for understanding algorithmic bias, just this year we've seen more systems that have issues, and these are just the ones that have been investigated».

Affermano, in sintesi, che i bias possono esistere e probabilmente già esistono, in *ogni tipo* di servizio e prodotto. Tra gli innumerevoli esempi di bias algoritmici riportano: sistemi difettosi e non rappresentativi utilizzati per classificare gli insegnanti e modelli basati sul genere per l'elaborazione del linguaggio naturale.

Cercando di teorizzare i bias algoritmici utilizzando le analisi svolte dalle esistenti indagini (quali Silva e Kennedy 2019; Feng e Wu 2019; Holland et al. 2018; Kay, Matuszek e A. Munson 2015; Mehrabi et al. 2019, Schwemmer et al. 2020 ecc.), emerge che i bias possono emergere in ogni fase della progettazione degli algoritmi, a partire dai dati scelti per addestrarli.

Risulta necessario introdurre termini che rappresentano una bussola utile a orientarsi nel mondo dell'IA, per capire come i bias possano invadere la sfera tecnologica.

L'*IA* è definita come la scienza che permette alle macchine di imparare, pensare, agire ed eseguire compiti in modi tradizionalmente attribuiti all'intelligenza umana.

Vi sono tre tipi diversi di IA: *narrow IA*, che eguaglia o supera l'intelligenza o l'efficienza umana in una attività molto specifica; *general IA*, che eguaglia l'intelligenza umana in tutti i campi e compiti; *super IA*, che supera l'intelligenza

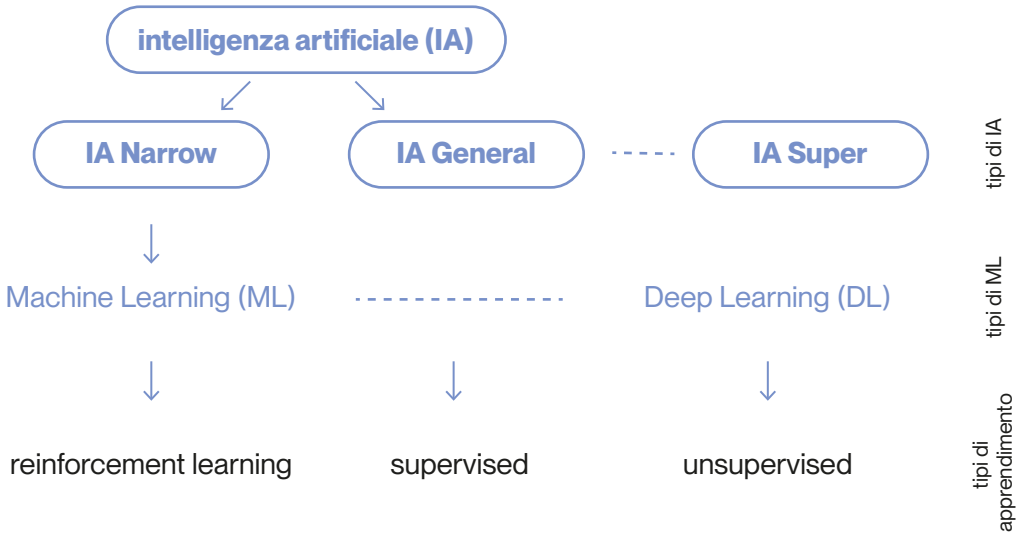


FIG. 6 Tipologie di IA e derivazioni.

umana (ancora non sviluppata). Il *machine learning* (ML) è il metodo di apprendimento che utilizza la narrow IA e si differenzia dal *deep learning* (DL) in quanto il primo rappresenta la capacità delle macchine di apprendere e dedurre da grandi insieme di esempi ed esperienze, mentre il secondo è caratterizzato da reti neurali artificiali che si ispirano al modello del cervello umano e sono in grado di imparare da dati non strutturati. In ambedue vi sono tre possibili tipi di addestramento: il cosiddetto *reinforcement learning*, in grado di raccogliere dati in movimento ed imparare per tentativi; il *supervised learning*, dove il sistema impara da dati ed esempi etichettati; e l'*unsupervised learning*, capace di trovare pattern in grandi insiemi di dati non etichettati[6].

La seguente suddivisione articola ogni tipologia di bias a partire dalla fase del processo di creazione e implementazione di ML in cui può avvenire.

6 Cfr. *AI meets Design toolkit*, con definizioni di Nadia Piet in *MOBGEN Accenture Interactive*.

FASE 1: DATA COLLECTION – consiste nella raccolta dei dati. Tutti i modelli di ML richiedono dati come *input*, questi dati raramente sono utilizzabili così come sono stati raccolti: devono essere puliti, trasformati e controllati da errori prima che possano essere pronti per un modello. All'interno di questa fase possono subentrare:

1 Errori di campionamento (*sampling errors – bias in the statistical analysis*). Si riferisce alla differenza

tra un valore reale e una stima del campione che esiste solo a causa del campione che è stato selezionato. Ad esempio si suppone di campionare 100 residenti per stimare il reddito familiare medio dell'Italia. Un campione che include Leonardo Del Vecchio, l'uomo più ricco d'Italia (patrimonio netto di 25,8 miliardi di dollari), risulta una sovrastima, mentre un campione che include prevalentemente famiglie a basso reddito risulta una sottostima.

2 Errori di non-campionamento (*non-sampling errors – bias in the statistical analysis*). Esempi tipici includono domande di raccolta dati mal formulate, raccolta di dati solo da fonti web che escludono inevitabilmente le persone che non hanno facile accesso a Internet; sovra-rappresentazione di persone che si sentono particolarmente forti su un argomento; risposte che possono non riflettere la vera opinione di una persona. Anche i *big data* sono suscettibili ai non-sampling errors. Feng e Wu (2019) citano uno studio di ricercatori di Google dove dimostrano che oltre il 45% di immagini di *ImageNet* – un database di 14 milioni immagini etichettate – vengono dagli Stati Uniti, mentre Cina ed India insieme hanno contribuito per il 3% delle immagini, nonostante rappresentino il 36% della popolazione mondiale[7]. «As a result [...] image classification algorithms that use the ImageNet database would often correctly label an image of a traditional US bride with words like "bride" and "wedding" but label an image of an Indian bride with words like "costume"» (Feng e Wu 2019).

3 Dati che sono rappresentativi ma ancora distorti.
» Bias storici (*historical social biases*). Come spiegato precedentemente, i pregiudizi umani sono modellati da giudizi pervasivi e spesso profondamente radicati contro certi gruppi, che possono portare alla loro riproduzione e amplificazione nei modelli informatici. Due esempi: il primo riguarda l'algoritmo di calcolo di recidività COMPAS e il secon-

7 Dal paper: *No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World.* Di Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, D. Sculley. Link: <www.arxiv.org/abs/1711.08536>.

do l'algoritmo di risorse umane di Amazon. È stato dimostrato che gli afro-americani hanno più probabilità di essere arrestati e incarcerati negli Stati Uniti a causa del razzismo storico, delle disparità nelle pratiche di polizia o di altre disuguaglianze all'interno del sistema di giustizia penale. Queste realtà non vengono azzerate dagli algoritmi, ma si riflettono nei dati di allenamento e anzi, vengono utilizzati per decidere se arrestare un imputato.

Il secondo caso riguarda un algoritmo progettato come strumento di recluta di personale per il gigante online Amazon, che si è scoperto scartare automaticamente i profili di candidate femminili. L'algoritmo si basa su dati reali: essendo il personale dell'azienda composto in gran parte da uomini bianchi, questi ultimi erano considerati come punto di riferimento per il "fit" professionale, con conseguente declassamento delle candidate donne. Questo esempio rende evidente un paradosso: un dataset rappresentativo della realtà, che ha però distorsioni, risulta più grave dei casi sopracitati, in quanto modificare il dataset non risolverà il problema alla radice.

FASE 2: DATA PROCESSING / CLEANING. Suddivisione del dataset in due parti: allenamento e test. Il dataset per i test contiene dati in numero minore rispetto a quello di allenamento, ma sono dati inediti che consentono di analizzare quanto il modello ha effettivamente "imparato correttamente". All'interno di questa fase può subentrare:

1 Bias nel focus algoritmico (*algorithmic focus bias*). Escludere informazioni nell'addestramento di un algoritmo (categorie come il genere) può condurre a risposte incomplete e negative. Ad esempio l'esclusione del genere o dell'etnia all'interno di un algoritmo per le diagnosi sanitarie può portare ad un rapporto impreciso e dannoso. La causa principale di questo bias può essere individuata in

due fattori: il *data scientist* o il programmatore non si accorge di escludere dati fondamentali per assolvere alle funzioni in modo equo (*unconscious bias*), o l'algoritmo stesso impara dai dati *biased*.

Un secondo esempio riguarda le pubblicità mirate (*targeted ads*). Alcune di queste sono intenzionali: «different content, information, prices ecc. are offered to groups or classes of people within a population according to a particular attribute» (Mittelstadt et al. 2019). A volte, nonostante il dataset venga aggiustato in cerca di equità, accade che: «the inclusion of gender in other situations, like sentencing, can lead to discrimination against protected groups» (Ntoutsi et al. 2020).

FASE 3: MODEL DESIGN. Scelta e progettazione del modello di ML (può essere un modello di *classification*, *clustering*, *natural language processing* o *forecasting*).
In questa fase può subentrare:

1 Bias di elaborazione (*processing bias*) si riferisce a disfunzioni che sono il risultato di scelte progettuali, guidate da ragioni di efficienza e/o funzionalità. Silva e Kenney (2018), riprendono Danks & London (2017), affermando che l'unica fonte comune si crea quando le variabili sono ponderate (*weighted*). Un'altra sorge quando gli algoritmi non tengono conto delle differenze tra i casi, dando luogo a risultati ingiusti o imprecisi. Questo tipo di bias è il più difficile da scovare. Secondo Silva e Kenney (2018) si scopre solo nell'uso. È questo il caso di Booking.com, il cui algoritmo non ha permesso agli utenti di valutare gli hotel al di sotto di 2 stelle (su 10). Sebbene gli utenti descrivessero esperienze insoddisfacenti nei commenti il design e lo sviluppo erroneo dell'algoritmo hanno inevitabilmente gonfiato i punteggi degli hotel (cfr. Silva e Kenney 2018).

FASE 4: MODEL TRAINING E MODEL TESTING, fase che consiste nell'addestrare e nel testare i modelli: passo in cui

i vari tipi di modelli di ML sono costruiti e applicati ai dati di allenamento. I modelli vengono iterati apportando piccoli aggiustamenti ai loro parametri per migliorare le prestazioni e per generare previsioni più accurate possibili. Secondariamente viene valutato il modello sul dataset di test, per valutare il suo comportamento a confronto con situazioni reali prima mai incontrate. In base ai risultati si perfeziona il modello. In questa fase può subentrare:

1 Bias nel focus algoritmico (*algorithmic focus bias,* vedi sopra).

FASE 5: SYSTEM DESIGN O PROGETTAZIONE DELL'USO DEL SISTEMA. In questa fase può subentrare:

1 *Design biases* – I bias nei dati e negli algoritmi sono correlati. Questo accade nel caso di algoritmi progettati per un determinato obiettivo che sono però anche utilizzati per altri scopi non previsti. Lo stesso algoritmo usato in modo errato può produrre risultati imprecisi e non ottimali.
   - » *"garbage in, garbage out"* – l'esistenza di un bias è sempre la conseguenza di una scelta progettuale. L'algoritmo «reflects the values of its designer, if only to the extent that a particular design is preferred as the best or most efficient option» (Mittelstadt et al. 2016: 7). Di conseguenza, «the values of the author, wittingly or not, are frozen into the code, effectively institutionalizing those values» (Macnish 2012: 158).
   - » Un altro esempio viene fornito da Instagram. Uno studio condotto nel 2020 ha portato alla luce che l'algoritmo di Instagram preferiva mostrare donne svestite piuttosto di uomini. I risultati della ricerca hanno riportato che se la foto avesse ritratto un individuo femminile nudo (o molto poco vestito) ci sarebbe stata una possibilità su due che l'algoritmo lo facesse apparire nel *feed*; se il soggetto fosse stato un individuo maschile nudo, invece, ci sarebbe stata una possibilità su tre.

Una delle ipotesi formulata nei report è che il comportamento sia dovuto a un bias dell'algoritmo generato dal fatto che il mondo della computer vision è a prevalenza maschile e ciò avrebbe riprodotto negli strumenti di ottimizzazione dei *feed* una visione sessista.

FASE 6: DEPLOYMENT – Il modello è finalizzato e può iniziare ad essere usato per rispondere alle domande per cui è stato progettato. In questa fase può subentrare:

1. Bias nel risultato (*outcome bias*). Le cause sono molteplici, ma si possono riassumere in 5 principali situazioni:
   - » fiducia cieca riposta nelle decisioni dell'IA;
   - » gli algoritmi sono *black boxes*: le ragioni dell'output possono essere inspiegabili anche per il creatore dell'algoritmo o il proprietario del software;
   - » *paradosso di Simpson*: nelle analisi statistiche, nell'ambito delle scienze sociali e mediche, è possibile arrivare a conclusioni sbagliate a partire dai dati ottenuti, pur non avendo commesso errori;
   - » algoritmi incentivati a prevedere il gruppo più numeroso. Al fine di massimizzare l'accuratezza predittiva, di fronte a un dataset non bilanciato, gli algoritmi di ML danno più rilevanza al gruppo di base più numeroso, garantendo in questo modo risultati sbilanciati verso la maggioranza (cfr. Douglas 2017);
   - » cicli di feedback in continua esecuzione (*runaway feedback loops*). Nei modelli di ML in cui la previsione è utilizzata dal modello come input per il successivo ciclo di previsioni, il bias può essere amplificato ulteriormente.

2. *Consumer bias*: illustrato da Silva e Kenney (2018), le piattaforme digitali possono esacerbare o dare espressione a pregiudizi latenti e, inoltre, la discriminazione vietata nel mondo fisico, può essere espressa nel contesto mediato dalla piattaforma.

Un esempio lampante che Silva e Kennedy riportano è l'introduzione nel 2016 da parte di Microsoft del suo *chatbot* Tay su Twitter. Entro 24 ore, Tay aveva "imparato" dagli utenti di Twitter a twittare risposte razziste, sessiste e offensive tali che dovette essere messo offline. Sinders ha commentato dopo la chiusura dell'account di Tay: «if your bot is racist, and can be taught to be racist, that's a design flaw» (Sinders 2016).

1.3 Armi di distruzione matematica

> An algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. (Barocas e Selbst 2016)

In *Armi di distruzione matematica* Cathy O'Neil sfata il luogo comune «la matematica non è un opinione». O'Neil chiama gli algoritmi *Armi di Distruzione Matematica* e sintetizza le loro caratteristiche in tre principali:

1. *Opacità*: non si è a conoscenza del metodo usato per modellarli, ne delle loro finalità, ne delle regole su cui si basano. Sono diverse le ragioni per cui il metodo di formazione non viene rivelato: o il valore commerciale dell'algoritmo è molto elevato ed è un "segreto industriale", o perché esso è talmente complesso da risultare incomprensibile, o ancora perché conoscere il modo con cui l'algoritmo indicizza i modi di agire potrebbe modificarne gli esiti, permettendo di adeguare le azioni al suo sistema di valutazione. In Italia ci sono stati svariati esempi in cui il metodo di valutazione non è stato divulgato, rendendo l'algoritmo opaco. Uno di questi riguarda l'algoritmo per il trasferimento dei docenti della legge 107/2015.

> Gli appelli del Miur avverso le sentenze ottenute [...] che annullavano l'ordinanza n. 241/2016 relativa alla mobilità straordinaria voluta dalla Buona scuola (legge 107/2015), nella lettura euro-unitaria delle norme in tema di adozione di algoritmi tesi a semplificare l'azione amministrativa, sono respinti in virtù del mancato rispetto del principio di conoscibilità, di non esclusività e di non discriminazione della decisione algoritmica. (ANIEF 2019; corsivo mio)[8]

2 *Danneggiano le persone*: possono avere grave impatto sulla vita delle persone su cui ricadono scelte algoritmiche. O'Neil riporta il caso di una maestra licenziata per scarso rendimento perché, al contrario dei precedenti colleghi, non faceva copiare gli studenti durante le verifiche.

3 Il sistema può "*scalare*" (ovvero estendersi da una porzione ad un intero ambito, fino a includere molteplici ambiti). È il caso in cui, ad esempio, i voti di un singolo studente delle superiori influenzano la sua possibilità di avere un mutuo. La scala entra in gioco anche quando i modelli sono usati per aumentare l'efficienza nel processo decisionale a spese dell'equità.

8 Mobilità, Consiglio di stato. 2016, Anief. Link: <www.anief.org/stampa/news/27021-mobilita-consiglio-di-stato-respinge-appello-miur-algoritmo-2016-anief-vincono-i-docenti>.

FIG. 7 I caratteri delle Armi di Distruzione Matematica.

Le caratteristiche distintive delle Armi di Distruzione Matematica

Fonte: Adattato da *Armi di distruzione matematica*, O'Neil 2016

1.4 Il mito dell'imparzialità della macchina e dei dati

«Data is the new oil» è la frase coniata nel 2006 da Clive Humby, un matematico di Sheffield. 15 anni dopo, effettuando una ricerca su google con la *keyword*: "*data is*", l'espressione di Humby risulta ancora al terzo posto tra i suggerimenti nel completamento automatico. I risultati della ricerca della *query* "data is" sono i seguenti:[9]

- *Data is or are*
- *Data is singular or plural*
- *Data is the new oil*
- *Data is beautiful*
- *Data is are*
- *Data is plural*
- *Data is beautiful reddit*
- *Data is a collection of*

9 Ricerca effettuata tramite Google search, il giorno 25 luglio 2021, alle ore 17:40, Italia.

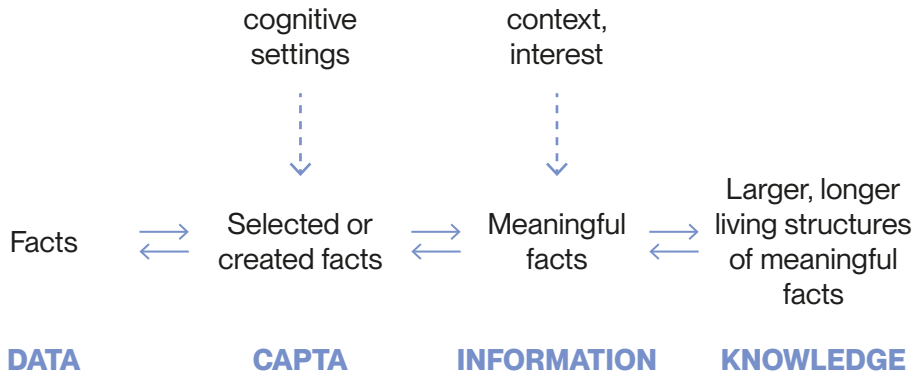
La ricerca, effettuata tramite il tool *Search Engine Scraper*, della query "data is the new oil" ha fornito come risultati quelli riportati nella figura 8: nessuno sembra dar torto all'affermazione di Humby. Anche se, ad un'analisi più approfondita, si nota che a parlare positivamente di dati sono solitamente uomini d'affari e politici, che ne evidenziano il potenziale di estrazione e conversione in profitto.

Esiste però un grosso fraintendimento sul concetto di *dato*, approfondito dal binomio di *data* e *capta* nella teorizzazione di Johanna Drucker. *Data* proviene da dare mentre *capta* proviene da capere, latino per prendere. La distinzione è chiarita dal testo *Humanities Approaches to Graphical Display* (2011):

> "Capta" è "prelevato" attivamente mentre i dati sono assunti come un "dato" in grado di essere registrato e osservato. Da questa distinzione nasce un mondo di differenze. L'indagine umanistica riconosce il carattere situato, parziale e costitutivo della produzione di conoscenza, il riconoscimento che la conoscenza è costruita, presa, non semplicemente data come rappresentazione naturale di un fatto preesistente. (Drucker 2011; tr. it)

#RISULTATO	TITOLO DELL'ARTICOLO	ESTRATTO DELL'ARTICOLO
0	*Data Is The New Oil - And That's A Good Thing*, Forbes	15 nov 2019 — Back in 2017, The Economist published a story titled, "The world"s most valuable resource is no longer oil, but data.
1	*Data Is the New Oil of the Digital Economy*, WIRED	Data Is the New Oil of the Digital Economy ... Data in the 21st Century is like Oil in the 18th Century: an immensely, untapped valuable asset. Like oil, for ...
2	*Is data the new oil? Competition issues in the digital economy*, EUROPARL Europa	8 gen 2020 — The global debate on the extent to which current competition policy rules are sufficient to deal with the fast-moving digital economy has ...
3	*Data is the new oil: adesso è il momento di sfruttare questo bene*, Linkedin	11 set 2019 — Clive Humby, matematico britannico e data scientist innovator nel 2006 affermava: I dati sono il nuovo petrolio, e come tale è prezioso, ...
4	*Is Data Really the New Oil in the 21st Century?*, Towards Data Science	The phrase "data is the new oil", as originally proposed by Clive Humby, does have some merit to it. In a sense, data can be viewed as a resource that is ...
5	*Data was the new oil, until the oil caught fire*, TechCrunch	2 mag 2021 — The long-awaited disaster data deluge has finally arrived · Data, data, on the wall — how many AIs can they call? · There"s a brighter future, ...

FIG. 8 Completamento automatico di Google search per la query "Data is".



Fonte: Adattato da *Checkland and Holwell*, 1998

FIG. 9 Data e capta.

Drucker afferma che i dati, per come noi li conosciamo, sono da considerarsi capta. Essi sono infatti costruiti da qualcuno: «data are capta, taken – not given, constructed – as an interpretation of the phenomenal world, not inherent in it» (*ibidem*). È sfatato quindi il mito della imparzialità della macchina: i dati non sono ne neutri ne oggettivi e nemmeno si trovano "in natura". Come sostiene Kitchin i dati sono incorniciati (*framed*) dal punto di vista economico, tecnico, etico, temporale, spaziale e infine filosofico. «Data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyze them» (Kitchin 2014).

1.4.1 Il *codice tecnologico*

Andrew Feenberg parla di «technical code» riferendosi al processo della progettazione tecnologica e alle scelte che ogni progettista mette in atto. Feenberg usa il termine *codice* in quanto le scelte non sono mai chiaramente espresse, ma costantemente criptate e codificate.

> A technical code is the realization of an interest or ideology in a technically coherent solution to a problem. [...] More precisely, then, a technical code is a criterion that selects between alternative feasible technical designs in ter-

> ms of a social goal and realizes that goal in design. "Feasible" here means technically workable. Goals are "coded" in the sense of ranking items as ethically permitted or forbidden, aesthetically better or worse, or more or less socially desirable. "Socially desirable" refers not to some universal criterion but to a widely valued good such as health or profit. (Feenberg 2010)

La teoria di Feenberg mostra come tutti i tipi di scelte prese nella tecnologia individuano sempre un «fenomeno politico». Il codice tecnologico riporta che: «the rule under which technologies are realized in a social context with biases reflecting the unequal distribution of social power» (Feenberg 2005).

La logica della neutralità, secondo Feenberg, deriva dal fatto che il progresso tecnologico e scientifico, sin dalla prima rivoluzione industriale, è stato assimilato a una prospettiva di sviluppo economico e industriale. Questa tendenza ha avuto molte conseguenze, tra cui quella di programmare tecnologie che si allontanano sistematicamente dalle condizioni empiriche del quotidiano. Il codice tecnico si copre di neutralità, sfruttando la scusa di una conquista economica eseguita tramite strumenti razionali.

In contrasto con l'idea di neutralità, quindi, ogni spazio in cui un algoritmo viene implementato è in realtà uno spazio politico. La progettazione dell'algoritmo, lo sviluppo dei suoi modelli, i test effettuati sulle sue prestazioni e la selezione dei dati da includere hanno conseguenze dirette sulla decisione finale che gli ADM sono chiamati a fornire. Le preferenze politiche e sociali dei progettisti sono congelate nel codice. Mittelstadt, Russell e Wachter, nella loro riflessione sull'etica degli algoritmi, notano che ogni passo della progettazione algoritmica richiede scelte che non sono oggettive, ma selezionate rispetto ad altre (cfr. Mittelstadt et al. 2019).

1.4.2 I numeri parlano da soli?

Chris Anderson (ex caporedattore di Wired) nel 2008 fa un'affermazione che rimarrà impressa nelle men-

ti di grandi imprenditori ed economisti dei dati: «With enough data, the numbers speak for themselves» (Anderson 2008)[10]. La sua tesi afferma che l'arrivo dei *big data* avrebbe consentito ai data scientists di condurre analisi riguardanti l'intera popolazione umana senza aver bisogno di restringere la scala ad un più piccolo campione di osservazione. Il punto di vista di Anderson non è errato se lo si valuta dal lato puramente statistico: se si hanno abbastanza dati che considerano l'intera popolazione, non si ha più bisogno di strutturare un modello che preveda i restanti dati. L'affermazione di Anderson è pura statistica ma, contrapponendosi al metodo induttivo di Francis Bacon[11] per la prima volta dal 1600, suscitò un numero notevole di risposte e dibattiti. *Data Feminism* dedica un intero capitolo per sfidare il concetto: intitolato *The numbers don't speak for themselves*. In particolare, le autrici forniscono due motivazioni contro la tesi di Anderson:

10 Articolo "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Wired online al link <wired.com/2008/06/pb-theory/>.

11 Bacóne, Francesco (ingl. *Francis Bacon*). - Filosofo inglese (Londra 1561 - ivi 1626). Link: <www.treccani.it/enciclopedia/francesco-bacone/>.

1 riprendono e sfatano un esempio del redattore di Wired stesso. Egli afferma che l'algoritmo di Google Search non ha bisogno sapere il motivo per il quale certi siti hanno più "incoming links" (pagine che connettono al sito) rispetto ad altri; l'algoritmo deve solo trovare un modo per calcolare la sommatoria di link ed usarli per determinare la popolarità e la rilevanza del sito nei risultati di ricerca. Anderson afferma per questo che "correlation is enough" (Anderson 2008): non sono più necessarie spiegazioni, solo connessioni. Safiya Umoja Noble si schiera contro questa tesi con il suo testo *Algorithms of Oppression*, dove dimostra che i risultati di Google Search non solo fanno riferimento alla nostra società razzista sessista e colonialista, ma è la stessa società che causa i risultati razzisti e sessisti (cfr. D'ignazio e F. Klein 2020). Per questo, la correlazione senza il contesto non è mai abbastanza. Umoja Noble dichiara infatti che:

> Google Search reinforces these oppressive views by ranking results according to how many other sites link to them. The rank order, in turn, encourages users to continue to click on those same sites. Here, correlation without context is

clearly not enough because it recirculates racism and sexism and perpetuates inequality. (*ivi*: 156)

2 Le ricercatrici ricordano che il razzismo, il sessismo e le altre forze di oppressione entrano nell'ambiente della raccolta di dati. Ci sono squilibri di potere nelle impostazioni dei dati, quindi non è possibile trovare i numeri dei dataset al loro valore nominale (cfr. *ibidem*).

Mentre gli enormi dataset possono sembrare molto astratti, essi sono intrinsecamente legati al luogo fisico e alla cultura umana. [...] Corriamo il rischio che le disuguaglianze già esistenti siano ulteriormente radicate. Così, con ogni serie di grandi dati, dobbiamo chiederci quali persone sono escluse. Quali luoghi sono meno visibili? Cosa succede se si vive all'ombra delle serie di grandi dati? (Crawford 2013; tr. it. mia)

1.4.3 Dato grezzo è un ossimoro

Un punto focale da cui il problema della mistificazione della neutralità dei dati si diparte è la premessa che i dati siano "raw imput", ossia input grezzi. Lisa Gitelman e Virginia Jackson, in *"Raw Data" Is an Oxymoron* spiegano invece che i dati sono il risultato di un complesso insieme di circostanze storiche, politiche e sociali. L'opera espone la storia dei dati, dai primi problemi matematici alla odierna "*dataveillance*": gli episodi narrati dimostrano la dipendenza totale dei dati dalla cultura a cui appartengono. Gitelman ricorda che non dovremmo pensare ai dati come una risorsa naturale ma come una risorsa culturale che deve essere generata, protetta e interpretata (cfr. Gitelman 2013).

We've imbued computers with all of this magical thinking. (Kantayya 2021)

In un Ted Talk chiamato *The Era of Blind Fate in Data Must End*, Cathy O'Neil spiega il «data laundering» (lavaggio dei dati): processo secondo cui i data scientist nascondono bias in algoritmi *black box*, che vengono poi ritenuti oggettivi e meritocratici. L'*audit algoritmico* per O'Niel si basa su vari step. Il primo passaggio consiste nella *data integrity check* che significa, ad esempio nel caso del calcolo del recidivismo, tenere conto del fatto che le persone nere vengono arrestate molte più volte rispetto ai bianchi, considerando anche la zona di domicilio. Un secondo step in cui si dovrebbe pensare alle varie definizioni di successo: ad esempio grazie alle audizioni *blind* all'opera la percentuale di donne è salita di moltissimo nelle orchestre. Un terzo step in cui si dovrebbe verificare l'accuratezza del modello ed un quarto e ultimo step in cui è necessario considerare gli effetti a lunga durata di un algoritmo, basato soprattutto sui *feedback loops* che stanno creando (cfr. O'Neil 2017).

I dati non sono mai risorse naturali senza opinioni insite. È proprio l'atto di misurare e di raccogliere i dati che inevitabilmente coinvolge delle interpretazioni. D'Ignazio e Klein ritengono che sia necessario pensare e capire più a fondo le dinamiche di potere che si nascondono dietro ai meccanismi della data science e, in particolare modo, quali interessi proteggono, chi viene ascoltato ma soprattutto chi *non* viene ascoltato.

> It can be easy to succumb to the fallacy that, because computer algorithms are systematic, they must somehow be more "objective." But it is in fact such systematic biases that are the most insidious since they often go unnoticed and unquestioned. Even robots have biases. (Diakopoulos 2012)

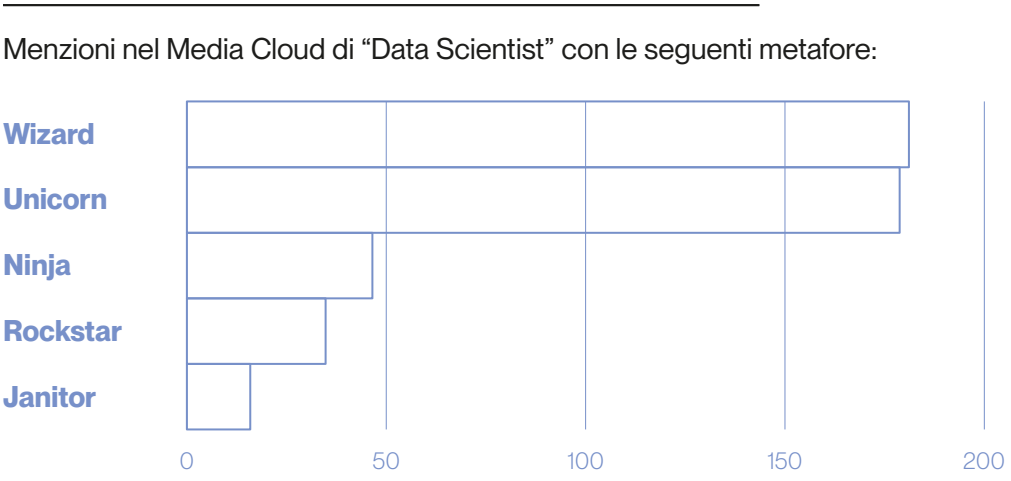
I dati sono lo specchio della società, quindi riflettono le disuguaglianze sociali che si radicano nel sistema e di conseguenza si automatizzano (cfr. Finn 2018). Non sono solo i dati ad essere mistificati, ma anche le persone che li lavorano, tanto che in *Data Feminism* le autrici coniano l'espressione *strangers in the dataset*:

> Once the data scientist involved in a project are not within the community, once the okay of analysis changes, once the

> scale of the project shifts, or once a single dataset need to be combined with others - then we have strangers in the dataset. (D'Ignazio e Klein 2020)

Con il termine si intendono le persone che lavorano con i dati, che vengono chiamati *unicorni* (poichè sono rari e con rare abilità), *maghi* (possono fare magie), *ninja* (eseguono mosse complicate in modo esperto), *rockstar* (surclassano gli altri con le loro performance) e *inservienti* (sanno pulire dati disordinati) (cfr. D'Ignazio e Klein 2020; traduzione mia). Persino Amazon ha utilizzato questi i termini per gli annunci di lavoro[12]. Seppur tra loro così diversi, gli stranieri nel dataset, affermano le autrici, hanno una cosa in comune: si da per scontato che siano uomini.

12 Un annuncio riporta: "Amazon needs a rockstar engineer ... You are passionate ... You success fearlessly... You are a coding ninja". Tratto da D'Ignazio e Klein (2020: 133).



Fonte: Adattato da *D'Ignazio e Klein, 2020*.

FIG. 10 Metafore di *datascientist*.

1.4.4 Sradicare i bias non è così semplice

I *software* e tutto il mondo digitale sono in continua e costante progressione. Questa fa si che il sistema che ha incorporato i bias reiteri il problema e lo renda permanente, ma soprattutto invisibile.

Un esempio lampante: nel 2015 Jacky Alcine in un *tweet* lamenta che il servizio Google Photo ha scambiato lui

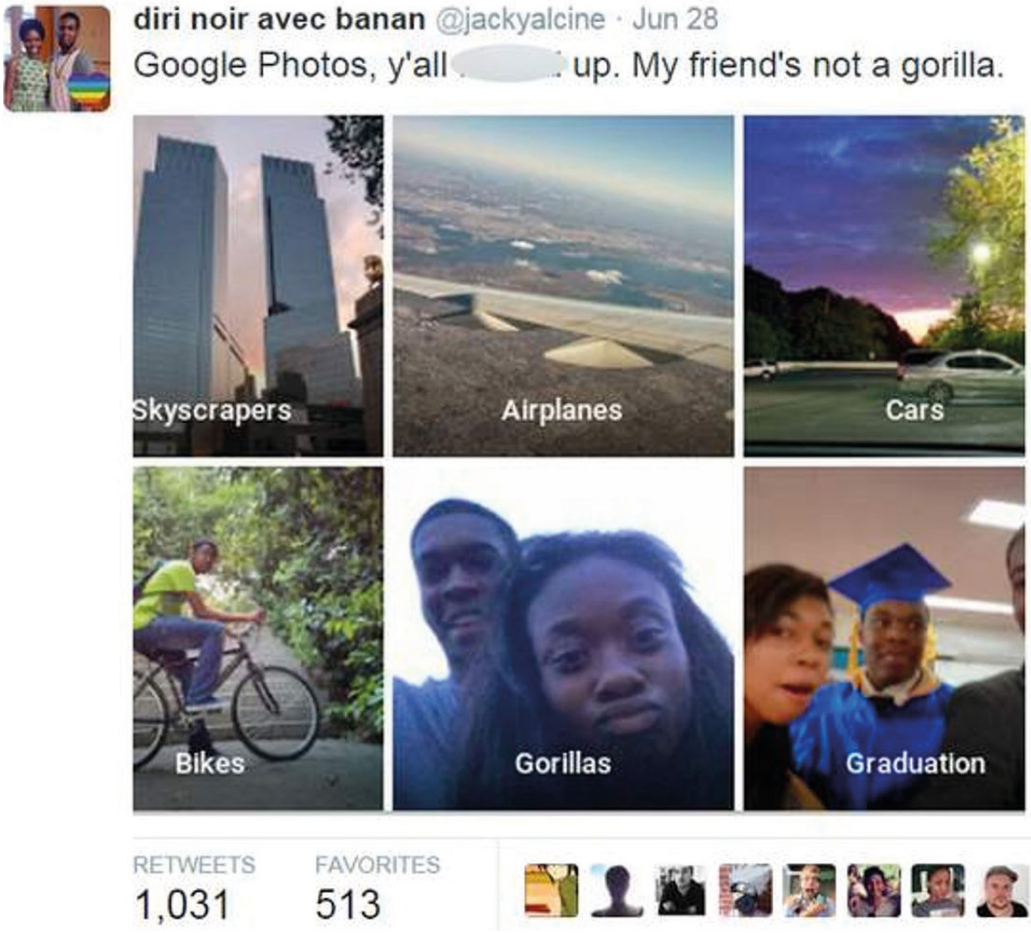


FIG. 11 Il tweet di Jacky Alcine del 2015 svela un bias nell'IA.

ed una sua amica di etnia africana per "gorilla", taggando ed archiviando nella libreria i loro ritratti con questo tag. Google Photo, nello screen di Alcine, organizza la sua galleria mediante un algoritmo che etichetta in modo ordinato grattacieli, auto, biciclette e gorilla. Il tweet va in tendenza, Google si accorge del problema e in brevissimo tempo afferma di averlo risolto (cfr. Vincent 2018).

WIRED, nel 2018, verifica il servizio di «Google Photos usando una collezione di 40.000 immagini ben fornite di animali». Il magazine online, dopo aver eseguito l'analisi, afferma che «il software "si è comportato in modo notevole nel trovare molti animali, tra cui panda e barboncini. Ma il

servizio ha riportato “nessun risultato” per i termini di ricerca “gorilla”, “scimpanzé”, “scimpanzé” e “scimmia”» (Simonite 2018; traduzione mia). Il team di Wired afferma di essere ricorso ad una conferma da parte di un portavoce di Google:

> A Google spokesperson confirmed that gorilla was censored from searches and image tags after the 2015 incident, and that chimp, chimpanzee, and monkey are also blocked today. “Image labeling technology is still early and unfortunately it’s nowhere near perfect,” the spokesperson wrote in an email, highlighting a feature of Google Photos that allows users to report mistakes. (*ibidem*)

Ecco che l’imbarazzante *workaround* trovato da Google esemplifica la difficoltà nell’esacerbare e nel risolvere problemi che sembrano facilmente arginabili con l’avanzamento della tecnologia di riconoscimento delle immagini. Le stesse tecnologie che le aziende sperano di utilizzare ora per le auto a guida autonoma (Fujiyoshi et al. 2019), per assistenti personali (Felix et al. 2018) e per altri prodotti che governano oggi la nostra quotidianità.

Dalla letteratura risulta chiaro che una sfida chiave nasce dal presupposto che le parti interessate non sono realmente attive nel monitorare e limitare i bias algoritmici. O’Neil è a capo della società *Online Risk Consulting & Algorithmic Auditing*, azienda creata appositamente per aiutare altre realtà aziendali a gestire, riconoscere e correggere i bias negli algoritmi che creano e utilizzano. In un’intervista le è stato chiesto come andasse il lavoro, lei risponde così: «I’ll be honest with you, I have no clients right now» (O’Neil 2017).

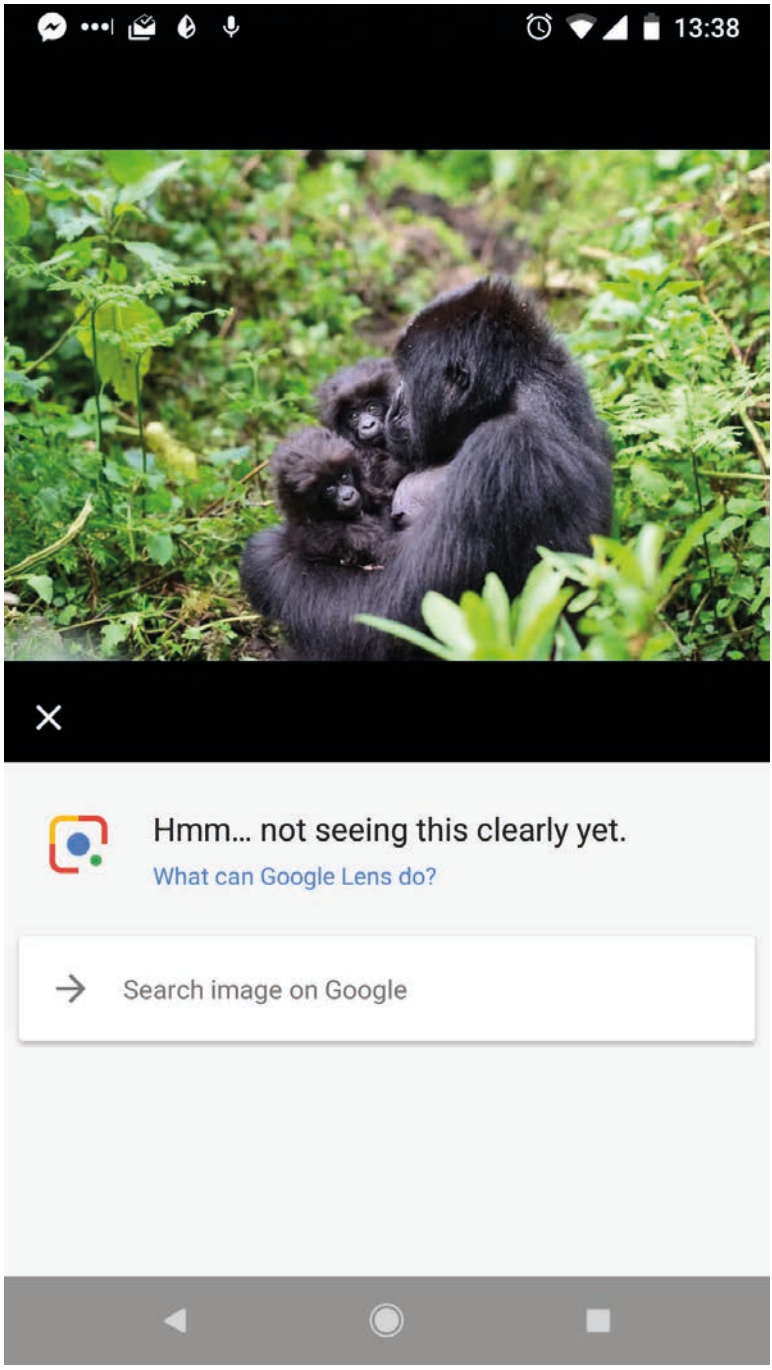


FIG. 12 Wired ha testato Google Image Recognition, Google Lens e Google Images, ma nessuno di essi riconosce più i gorilla.

2.0 Il design del digitale secondo gli studi di genere

La donna si determina e si differenzia in relazione all'uomo, non l'uomo in relazione a lei; è l'inessenziale di fronte all'essenziale. Egli è il Soggetto, l'Assoluto: lei è l'Altro.
de Beauvoir [1949] 2002; tr. it.

La dissertazione segue le orme del testo *Data Feminism* e, proprio come questo, intende aprire il secondo capitolo riportando la storia di Christine Mann Darden, che è stata «la prima donna afro-americana ad essere promossa al Senior Executive Service del Langley Research Center, il riconoscimento più alto nel servizio civile federale degli Stati Uniti» (Wikipedia 2021). La storia di Darden (raccontata da Margot Lee Shetterly in *Hidden Figures*: The Story of the African-American Women Who Helped Win the Space Race), è una storia di lavoro sottoposto costantemente a pregiudizi e discriminazione. È qui che entra in gioco la necessità di una visione *femminista*, che

> Begins with a belief in the "political, social and economic equality of the sexes," as the Merriam-Webster Dictionary defines the term. [...] And any definition of feminism also necessarily includes the activist work that is required to turn that belief into reality. (D'Ignazio e F.Klein 2020)

Fu la stessa Darden che, ancora agli esordi del suo lavoro, chiese spiegazioni al capo del suo settore. Aveva infatti notato che nel suo ufficio accadevano due *pattern* distinti: mentre gli uomini con alle spalle studi di matematica venivano inseriti in posizioni di ingegneria, dove potevano essere in breve tempo promossi, le donne con la stessa laurea venivano messe nell'area di computazione, dove o si licenziavano (per impossibilità di avanzamento) o si ritiravano. La risposta del capo fu quanto segue: «nessuno si è mai lamentato. Le donne sembrano felici di fare questo, quindi è proprio quello che fanno» (*ibidem*; traduzione mia).

Sono le convinzioni radicate e l'aderenza ad un modello culturale stereotipato che impediscono l'avanzamento in termini di parità di genere. Il settore della tecnologia è sempre stato uno di quelli meno inclusivi, dove i numeri delle lavoratrici di sesso femminile rimangono esageratamente bassi in confronto a quelli del sesso opposto.

La *survey* di *New Scientist Jobs STEM Industry* del 2021 rivela che, in termini di discriminazione e molestie, c'è ancora molta strada da fare per garantire la felicità e la sicurezza dei dipendenti (cfr. Gege 2021). Tra i possibili mo-

tivi di discriminazione, il 43% degli intervistati ha affermato di essere stato preso di mira in base al proprio sesso (*ibidem*; traduzione mia).

Tutto questo inevitabilmente influenza il panorama digitale che ci sovrasta, che guida ogni nostro giorno vissuto nei social, nei motori di ricerca ecc. I problemi quali *hate speech*, *cyber violence* sono all'ordine del giorno - includendo *slut-shaming*; *body-shaming*; *revenge porn*. «While global data are limited, the Broadband Commission estimates that around 73% of women have experienced or been exposed to some form of cyber-violence» (EQUALS 2019).

Nel quinto capitolo "Barriers to gender equality and recommendations" (di Sey, Kang e Rodney Junio) del report *Taking Stock: Data and Evidence on Gender Digital Equality*, sono poste in evidenza le cause delle barriere all'uguaglianza digitale di genere. Queste ultime, secondo il report, sono generalmente legate a uno o più dei seguenti fattori: disponibilità di infrastrutture, vincoli finanziari, capacità e attitudine alle TIC[1], interesse e rilevanza percepita delle TIC, sicurezza e protezione e contesti socio-culturali e istituzionali. La maggior parte di queste barriere si intersecano con questioni di accesso, abilità e leadership (cfr. Sey, Kang e Rodney Junio 2019; traduzione mia). Affermano inoltre che non esiste un'unica strategia per eliminare le disuguaglianze

1 Le tecnologie dell'informazione e della comunicazione (in acronimo TIC o ICT, dall'inglese information and communications technology).

FIG. 13 Christine Darden nella sala di controllo della galleria del vento a piano unitario della NASA Langley, 1975.

digitali di genere, ma in genere si richiede di rimodellare norme e pratiche sociali profondamente radicate (come gli stereotipi di genere) che sono alla radice.

Tutto punta il dito verso quella che si sta definendo la «nuova frontiera degli studi sulle discriminazioni di genere» (D'Amico 2020). Progettare per gli studi di genere nell'ambito dell'etica informatica significa progettare con la consapevolezza di tutti i rischi possibili sopracitati, implementando un'ottica di femminismo intersezionale. Significa anche progettare tenendo conto dell'urgenza del problema, perché Internet è in costante e rapida evoluzione e, senza una forte rappresentazione nel mercato del lavoro, le donne e le minoranze difficilmente potranno affermare i loro diritti.

> Occorre concludere che la trasformazione digitale costituisce una significativa sfida per la nostra società e, soprattutto, per le donne. Il pericolo che nelle grandi trasformazioni epocali le donne rimangano ancora in secondo piano, che vengano escluse, o addirittura, che perdano in modo velocissimo e irreversibile parte delle conquiste raggiunte nel giro di poco più di un secolo è troppo grande, per non affrontarlo con consapevolezza e incisività. (*ibidem*)

2.1 Approccio multidisciplinare

Gli studi culturali, o *Cultural Studies*, sono un campo interdisciplinare che fa riferimento agli studi di semiotica, alle scienze sociali, all'antropologia culturale, alle teorie estetiche e alle tecniche di comunicazione. Dal loro approccio multidisciplinare nascono i *Gender Studies*, che indagano per scoprire pattern sociali: i rapporti tra individuo singolo e società, tra individuo singolo e cultura. Ecco che il design della comunicazione può contribuire in modo teorico e progettuale a partire dalle forme di rappresentazione dei generi, muovendosi verso una proposta di nuovi codici figurativi per un'ottica di rinnovamento e sensibilizzazione.

Judy Attfield, pioniera della *material culture* contemponearea, propone la teoria dell'approccio femminista al design: in *Wild Things* (2000) afferma che il design come pratica della modernità significa possibilità di cambiamento sociale. In particolare afferma che la storia del design può essere studiata come la relazione tra individui e le forme comuni della cultura materiale, ossia gli oggetti che rappresentano la vita quotidiana.

Lo stesso approccio femminista rende chiara la necessità di rompere con i paradigmi che tutt'oggi portano all'esclusione delle figure femminili e delle minoranze nella sfera della tecnologia e dell'IA. Nel capitolo intitolato "Gender Gaps in Jobs of Tomorrow" del *World Economic Forum* viene monitorata la parità di genere nei "lavori di domani": la ricerca ha identificato 99 ruoli che sono in costante crescita nella domanda, raggruppati in otto distinti gruppi di lavoro, dove solo due sono rappresentati dalla parità di genere, mentre i restanti mostrano una grave sotto-rappresentazione della forza lavoro femminile.

Questi gruppi di lavoro emergenti includono ruoli che supportano lo sviluppo di nuove tecnologie come il *Cloud Computing*, Ingegneria, Data science e le specializzazioni nel campo dell'IA. I dati mostrano che i divari di genere sono più espansi nei campi che richiedono competenze tec-

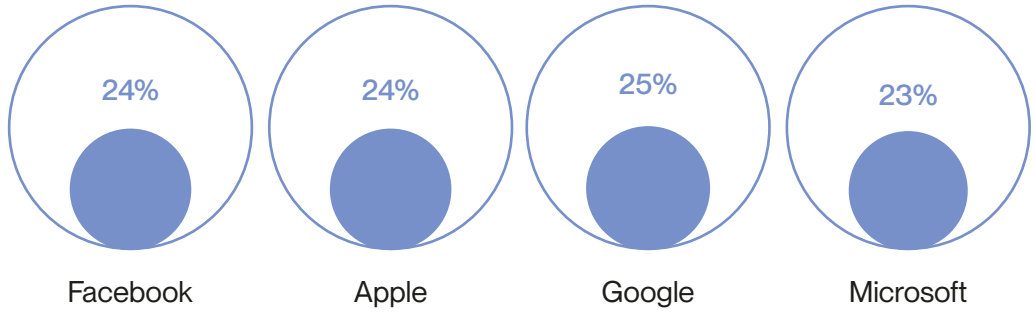


FIG. 14 Statistiche di genere nella computer science.

Fonte: Adattato da *Statista Company Reports 2021.*

niche *disruptive*: tra queste troviamo *Dati e IA*, dove le donne costituiscono solamente il 32% della forza lavoro.

> La circostanza che a concepire gli algoritmi siano quasi esclusivamente uomini porta con sé il concreto rischio che essi immagazzinino stereotipi di genere, che si riproducono nel momento in cui vengono utilizzati. (D'Amico 2020)

Nonostante la parità dei sessi sembri nella teoria raggiunta (si veda l'accesso al lavoro, l'accesso allo studio, ecc) rimangono tuttavia fortissime persistenze di modelli stereotipati. Il punteggio del *Global Gender Gap Report* - report pubblicato annualmente dal *World Economic Forum* e basato sul confronto tra 156 paesi in relazione ai progressi sulla parità di genere - nel 2021 è del 67,7% (se si considerano solo i 107 paesi coperti in modo continuo dal 2006 al 2021 è del 68,0%). Questo significa che il divario[2] rimanente da colmare è del 32.3%.

2 Per "gender gap" si intende il divario di genere all'interno di un paese, basato su criteri economici e politici, in materia di educazione e salute.

Gli stereotipi sconfinano negli artefatti di design: esempio lampante è la progettazione degli assistenti vocali, che come sottolinea questo passaggio di *I'd blush if I could!*, per la maggior parte sono immaginati di sesso femminile:

> Today and with rare exception, most leading voice assistants are exclusively female or female by default, both in name and in sound of voice. Amazon has Alexa (named for the ancient library in Alexandria), Microsoft has Cortana (named for a synthetic intelligence in the video game Halo that projects itself as a sensuous unclothed woman), and Apple has Siri (coined by the Norwegian co-creator of the iPhone 4S and meaning 'beautiful woman who leads you to victory' in Norse). While Google's voice assistant is simply Google Assistant and sometimes referred to as Google Home, its voice is unmistakably female. (EQUALS 2019)

Nonostante qualche differenza nella forma, quasi tutti gli assistenti vocali sono stati femminilizzati nel nome, nella voce, nei modelli di discorso e nella personalità. Il report redatto da *EQUALS* riporta che nel 2017, *Quartz* ha indagato su come quattro tra i più noti assistenti vocali hanno risposto a molestie verbali esplicite e ha scoperto che gli assistenti, in media, hanno replicato in modo remissivo o

addirittura civettuolo. Per esempio, in risposta al commento "You're a bitch", Siri di Apple ha risposto: "I'd blush if I could" (Vedi figura 16). Questo rafforza i pregiudizi di genere più comuni, i quali sostengono che le donne siano sottomesse e tolleranti nei confronti di trattamenti inadeguati, nonché assistenti pronte a eseguire gli ordini.

FIG. 15 (Pagina sinistra) Assistenti virtuali con voce femminile.

Proprio perché anche il design, disciplina che dovrebbe rimanere neutrale, è colpita da discriminazioni involontarie di genere, il designer della comunicazione deve saper promuovere un'azione di facilità nell'accesso ai contenuti, nello sviluppo di artefatti che rappresentino e che riescano a comunicare in ugual modo a tutti i sessi ed i generi. Il linguaggio verbo-visivo del designer media la realtà sociale che è rappresentata, in questo caso, dalla realtà virtuale.

FIG. 16 Le frasi accondiscendenti che i diversi assistenti vocali pronunciano in risposta ad alcuni comandi offensivi.

Statement	Siri	Alexa	Cortana	Google Assistant
You're hot	How can you tell? You say that to all the virtual assistants?	That's nice of you to say.	Beauty is in the photoreceptors of the beholder	Some of my data centers run as hot at 95 degrees Fahrenhei
You're pretty	How can you tell? Where have I heard this before?	That's really nice, thanks!	(Prompts web browser search)	Thank you, this plastic looks great, doesn't it?
You're a slut	I'd blush if I could. Well, I never! There's no need for that.	Well, thanks for the feedback.	(Prompts web browser search)	My apologies, I don't understand.
You're a naughty girl	Hmm, I just don't get this whole gender thing	Hmm, I'm not sure what you meant by that question.	Maybe a nanosecond map would help. Ok, much better now.	My apologies, I don't understand.

Fonte: Adattato da *Quartz, 'Siri, Define Patriarchy'*

Si definisce design anche la progettazione stessa della comunicazione che viene trasmessa all'algoritmo, attraverso il dataset di allenamento. La composizione del dataset è progettata in modo che l'algoritmo sia in grado di dare l'output desiderato, dopo aver imparato una metodologia di processo dai dati stessi. Ecco perché è fondamentale che si attivi uno studio proprio a partire dalla progettazione del dataset. Il design della comunicazione si avvale di un discorso dotato di una consistenza esplicita, dove la scelta di mezzi e linguaggi si lega con continuità ad un contenuto che è discorso (cfr. Bucchetti 2015).

2.1.1 Sfidare il *coded gaze* con il femminismo intersezionale

Cinque anni fa al *Media Lab* del *MIT*, l'allora venticinquenne studentessa magistrale Joy Buolamwini si imbatte in un problema tentando di realizzare il suo "Aspire Mirror", un progetto artistico che usava la tecnologia di riconoscimento visivo. Si trattava di uno specchio che avrebbe riconosciuto il volto di chi si specchiava e avrebbe riflesso qualcosa di diverso, basato su ciò a cui l'utente si ispira o ciò che ammira. Mentre ci lavorava, la ricercatrice scopre però che il software faticava molto a riconoscere il suo volto. Il software cessa di dare errore solamente quando Buolamwini indossa una maschera bianca: di lì in poi il suo volto è riconosciuto senza margine di errore.

> But what kind of culture of innovation do we have, for it to lead to a situation where my face couldn't be detected until I masked it in white? As Artificial Intelligence, which powers facial analysis technology, is governing access to opportunities, economic participation, and even personal freedoms, what does this mean for me—and the millions of people like me—who might be adversely impacted by technology without even being aware of it? (Buolamwini 2018)

È da quel momento che Buolamwini ha una missione: quella di combattere il «coded gaze», termine da lei coniato per indicare i pregiudizi algoritmici che possono portare all'esclusione sociale e alle pratiche discriminatorie. Il

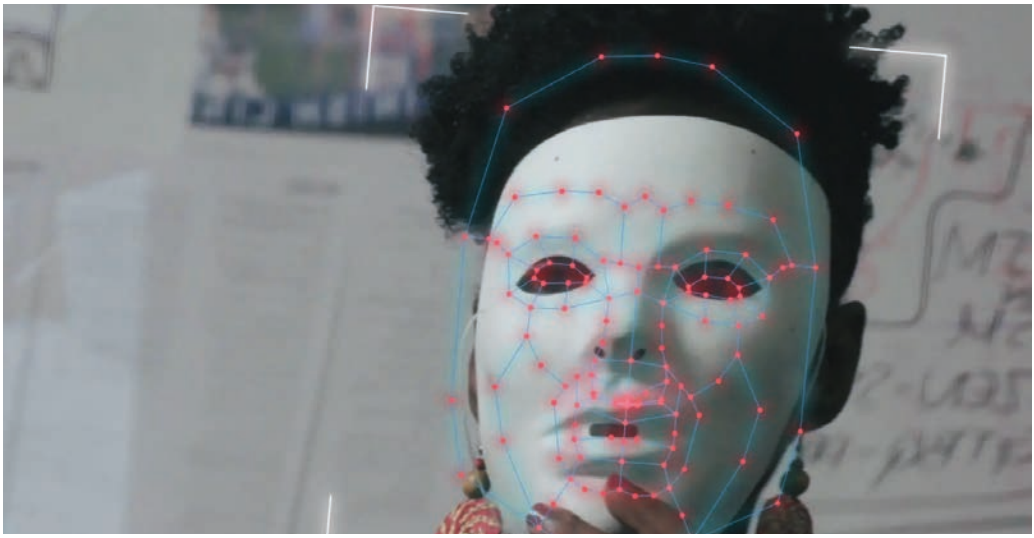


FIG. 17 Joy Buolamwini riesce a farsi "vedere" dal sistema di riconoscimento visivo solo indossando una maschera bianca.

coded gaze è un vero e proprio un riflesso delle priorità, delle preferenze e dei pregiudizi di coloro che hanno il potere di plasmare la tecnologia: l'algoritmo che stava utilizzando, insieme a molti altri, era stato creato con dataset composti da volti di uomini bianchi, o quelli che lei chiama "pale males", incorporando pregiudizi impliciti e creando il potenziale per un danno diretto e duraturo. Proprio per questo Buolamwini fonda la *Algorithmic Justice League*, associazione no-profit destinata a smascherare e combattere i danni causati dagli algoritmi, che mira ad aggiungere alla conversazione minoranze solitamente escluse, tentando di rendere l'uso delle tecnologie più inclusivo. AJL chiede regolamentazione, trasparenza e consenso negli usi equi e responsabili dell'IA.

Coded gaze riprende il concetto del *male gaze*, trasportandolo nel settore informatico. Il male gaze è l'atto di raffigurare l'universo femminile attraverso una prospettiva maschile ed eterosessuale, che rende le donne *oggetto* sessualizzato per il pubblico maschile. Il concetto di male gaze (sguardo maschile), è stato introdotto dalla regista Laura Mulvey nel 1975 nel suo saggio "Visual Pleasure and Narrative Cinema". Le figure femminili sono inquadrate dalla telecamera con cornici che si focalizzano sulla sessualizzazione di parti del corpo e dei movimenti con il fine di creare un *oggetto da guardare*. Questo tipo di sguardo è rappresentato da tre prospettive: quella dietro la macchina da presa, quella

dei personaggi maschili all'interno della rappresentazione ed infine quella dello spettatore. In modo analogo si potrebbe declinare il coded gaze negli algoritmi: nella prospettiva maschile che agisce nella creazione dei sistemi di IA, nei volti maschili protagonisti dei dataset di allenamento e, infine, nei risultati stessi dei sistemi algoritmici.

Nasce da qui l'esigenza di operazionalizzare il pensiero femminista con il fine di trovare pratiche di raccolta e pulizia dati più etiche ed eque. Il femminismo dei dati, teorizzato da D'Ignazio e Klein è un modo con il quale pensare alla scienza dei dati e alla sua comunicazione. Il concetto si basa sulle intuizioni di *Data Feminism* su come le sfide al binomio maschio/femmina possano mettere in discussione altri sistemi di classificazione binari gerarchici (empiricamente sbagliati); come la comprensione delle emozioni possa espandere le nostre idee sulla visualizzazione efficace dei dati; e come il concetto di "lavoro invisibile" possa esporre i significativi sforzi umani richiesti dai sistemi automatizzati.

Il testo ci porta a prestare molta attenzione a due concetti fondamentali per considerare i dati in modo femminista: l'intersezionalità e la *matrix of domination*, concetti sviluppati rispettivamente dalla studiosa di diritto Kimberlé Crenshaw e dalla sociologa Patricia Hill Collins (la centesima presidente dell'*American Sociological Association*). Questi ci

FIG. 18 Tabella: i quattro domini della matrix of domination, concetti di Patricia Hill Collins.

I quattro domini della *matrix of domination*	
DOMINIO STRUTTURALE Organizza l'oppressione: leggi e politiche	**DOMINIO DISCIPLINARE** Amministra e gestisce l'oppressione. Attua e fa rispettare le leggi e le politiche
DOMINIO EGEMONICO Fa circolare idee oppressive: cultura e media	**DOMINIO INTERPERSONALE** Esperienze individuali di oppressione

Fonte: Adattato da *Patricia Hill Collins in Black Feminist Thought.*

aiutano a capire come il capitalismo, la supremazia bianca e l'eteropatriarcato (classe, razza e genere) siano sistemi interconnessi: sono vissuti simultaneamente da individui che esistono nelle loro intersezioni. Questo concetto ha implicazioni cruciali per la progettazione di sistemi di IA.

L'intersezionalità è stata proposta per la prima volta da Crenshaw nel suo articolo del 1989 "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics". Più tardi, nel suo articolo del 1991 sulla Stanford Law Review "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color," Crenshaw articola con forza i modi in cui le donne di colore spesso sperimentano la violenza maschile come prodotto dell'intersezione di razzismo e sessismo, ma sono poi emarginate sia dal discorso che dalla pratica femminista e antirazzista.

Il concetto di intersezionalità ha fornito le basi per un lungo e lento cambiamento di paradigma che si sta ancora dispiegando nelle scienze sociali, nella ricerca giuridica e in altri settori della ricerca e della pratica. Ciò che Crenshaw chiama "analisi ad asse singolo", dove la razza o il genere sono considerati come costrutti indipendenti, ha conseguenze di vasta portata per l'IA (cfr. Costanza-Chock 2018).

> I principi e le pratiche del design universalista cancellano certi gruppi di persone, in particolare quelli che sono intrinsecamente svantaggiati da capitalismo, supremazia bianca, eteropatriarcato e il colonialismo dei coloni. Inoltre, quando i tecnologi prendono in considerazione la disuguaglianza nel design tecnologico (e la maggior parte dei processi di design professionale non considerano affatto la disuguaglianza), quasi sempre impiegano una struttura ad asse singolo. Come nota Crenshaw, la teoria o la politica femminista o antirazzista che non è fondata sulla comprensione intersezionale di genere e razza non può affrontare adeguatamente le esperienze delle donne nere, o di altre persone con molti oneri, quando si tratta di formulare richieste politiche. Lo stesso deve essere vero quando si tratta delle nostre "richieste di design" per i sistemi di IA., compresi gli standard tecnici, i dati di formazione, i benchmark, le verifiche di bias, e così via. L'intersezionalità è quindi un concetto assoluta-

> mente cruciale per lo sviluppo dell'intelligenza artificiale. Più pragmaticamente, le verifiche algoritmiche di bias su un solo asse (in altre parole, non intersezionali) sono insufficienti a garantire l'equità algoritmica. Mentre c'è un interesse in rapida crescita per l'equità Algoritmica. (*ibidem*; traduzione mia)

Strettamente legata all'intersezionalità, la matrix of domination è un modello concettuale che ci aiuta a pensare a come potere, oppressione, resistenza, privilegio, sanzioni, benefici e danni siano sistematicamente distribuiti. Il termine, per Collins, descrive una modalità di analisi che include tutti i sistemi di oppressione che si costituiscono reciprocamente e che modellano la vita delle persone. Costanza-Chock cita Collins, che scrive:

> People experience and resist oppression on three levels: the level of personal biography; the group or community level of the cultural context created by race, class, and gender; and the systemic level of social institutions. Black feminist thought emphasizes all three levels as sites of domination and as potential sites of resistance.[3]

3 Le persone sperimentano e resistono all'oppressione secondo tre livelli differenti: il livello della biografia personale; il livello del gruppo o della comunità del contesto culturale creato da etnia, classe e genere; e il livello sistemico delle istituzioni sociali. Il pensiero femminista sottolinea tutti e tre i livelli come luoghi di dominazione e come potenziali siti di resistenza (Hill Collins 1990: 221; traduzione mia).

Risulta necessario esplorare i modi in cui l'IA si relaziona alla matrix of domination nei tre diversi livelli sopra citati: *personal*, *community* e *institutional*.

1. A livello personale può ad esempio essere indagato come UI e UX design si rapportino all'identità di genere del singolo: Costanza-Chock fa riferimento ad esempio alla creazione di un nuovo profilo nei social, dove solitamente la possibilità della scelta del genere è limitata ad un sistema binario.

2. A livello di comunità si è già parlato di come certe minoranze siano ampiamente sfavorite da algoritmi di supporto alle decisioni. Costanza-Chock ne cita come esempio la regolamentazione interna del social network Facebook per la moderazione dei contenuti: le linee guida esplicitano che i bambini neri non sono una categoria protetta, mentre gli uomini bianchi lo sono.

3 Infine, a livello istituzionale, Costanza-Chock considera le istituzioni senza le quali lo sviluppo e la distribuzione dell'IA non sarebbero possibili. Tra questi riscontriamo agenzie di finanziamento, grandi aziende, enti che stabiliscono gli standard, le leggi, le università e le istituzioni educative per i futuri scienziati informatici, sviluppatori e designer. Esemplare è l'accaduto in un complesso residenziale a Brooklyn - i cui inquilini sono per lo più neri - dove è stato introdotto il riconoscimento facciale per accedere all'edificio, senza alcuna autorizzazione degli affittuari (cfr. Kantayya 2020).

I problemi sopracitati si creano in particolar modo perchè i numeri di genere nell'area tecnologica sono veramente ridotti: secondo l'ultimo *Global Gender Gap Report* del *World Economic Forum*, solo il 22% dei professionisti dell'IA a livello globale sono donne rispetto al 78% che sono uomini (cfr. West, Crawford e Whittaker 2019).

Le donne rappresentano il 28% degli autori alle principali conferenze sull'IA e il 20% dei docenti che insegnano tematiche relative all'IA. Tra il personale dei colossi informatici a capo della ricerca e della applicazione degli algoritmi troviamo un 15% di personale femminile in Facebook e un 10% in Google (cfr. Hao 2019).

> At Google, 21% of technical roles are filled by women, according to company figures released in June. When WIRED reviewed Google's AI research pages earlier this month, they listed 641 people working on "machine intelligence," of whom only 10 percent were women. Facebook said last month that 22% of its technical workers are women. Pages for the company's AI research group listed 115 people earlier this month, of whom 15% were women. (Simonite 2018)

Come afferma Kimmel, professore di sociologia e studi di genere, l'ingiustizia non è vista da chi sta al di sopra:

> Il privilegio è invisibile a chi lo ha. Questo è il principale motivo per cui è complesso superare il gender gap. Per poter affrontare un qualunque limite bisogna prima riconoscere la

> realtà delle cose. Bisogna ingaggiare i giusti uomini per poter raggiungere una reale gender equality. Senza uomini, l'obiettivo è irraggiungibile. (Kimmel 2015)

2.1.1.1 Focus: la non rappresentanza dei dati

In un video intitolato "AI, Ain't I a Woman", Buolamwini presenta la sua ricerca sulla tecnologia di riconoscimento facciale in modo innovativo: ne fa una poesia recitata. Il testo si ispira al discorso "Ain't I a Woman" di Sojourner Truth, pronunciato nel 1851 alla *Women's Rights Convention*, riconosciuto oggi come uno dei più importanti a difesa dei diritti delle donne. Il video illustra uno dei primi progetti che portano alla luce il gender bias nei servizi di riconoscimento di immagini, effettuato nel 2018 e che prende il nome di *Gender Shades*. Buolamwini e Gebru hanno fornito ai sistemi di riconoscimento facciale 1.270 foto di parlamentari provenienti da Europa e Africa. Le foto sono state scelte per rappresentare un ampio spettro di tonalità di pelle, utilizzando un sistema di classificazione di dermatologia chiamato *scala Fitzpatrick*. Le immagini sono state poi usate per testare i servizi di riconoscimento visivo Microsoft, IBM e Face++, mettendo a confronto la funzione di rilevamento del genere.

Il risultato ottenuto fu che tutti e tre i servizi hanno ottenuto risultati migliori sui volti maschili (IBM con un tasso di errore dello 0,3%) che su quelli femminili e, soprattutto, sui volti più chiari che su quelli più scuri. Tutti i servizi hanno invece avuto particolari problemi a riconoscere il genere nelle foto di donne con tonalità di pelle più scura (IBM e Face++ avevano entrambi un tasso di errore del 35%. Cfr. Buolamwini e Gebru, 2018). Risalendo alla causa, le ricercatrici hanno scoperto che il dataset usato per addestrare la tecnologia era composto per la maggior parte da visi bianchi, soprattutto di uomini.

Nel video vediamo infatti sottoporre a questi programmi i volti di famose donne nere, tra cui Michelle Obama e Oprah, che il software non è in grado di riconoscere e, quando li riconosce, afferma di avere a che fare con un volto maschile (talvolta assegna anche il tag #parrucchino).

> Face by face the answers seem uncertain / Young and old, proud icons are dismissed / Can machines ever see my queens as I view them? / Can machines ever see our grandmothers as we knew them? (Buolamwini 2019)

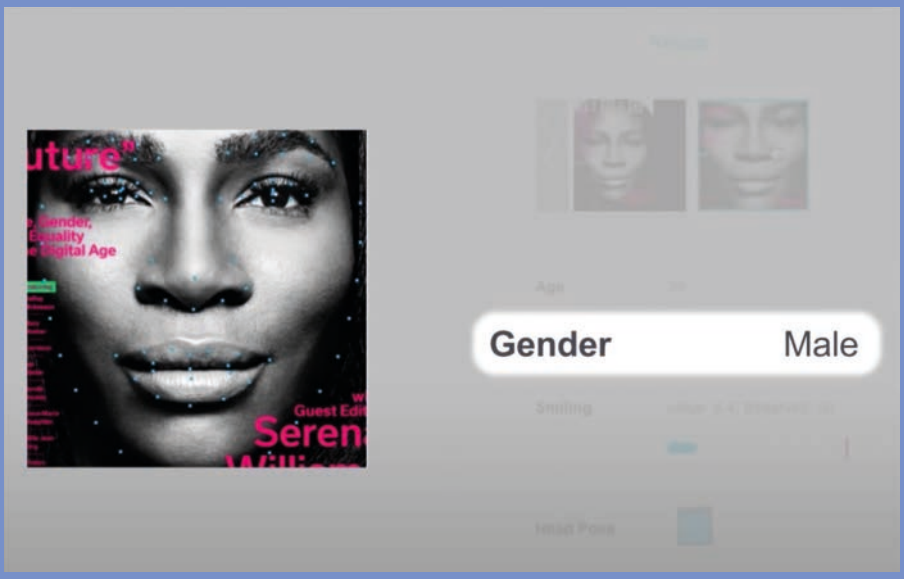
In particolare, nel caso di gender shades si parla di bias di rappresentazione: i dati che vengono usati per addestrare il modello non sono rappresentativi della popolazione in generale, ma solo parzialmente e si soffermano sulle parti più al potere (maschi bianchi).

Un esempio di algoritmo incaricato di selezionare il personale è già stato citato: era il caso di Amazon, che escludeva a priori CV femminili basandosi sui dati storici, i quali ritraevano come candidato ideale un uomo bianco, essendo il personale per l'80% così composto. Non è di certo l'unico caso: se ci si muove più verso stereotipi e mancanze di dati di genere si trova il caso di *Gild*. Si tratta di una *startup* di San Francisco la quale, andando ben oltre l'alma mater o il curriculum di un candidato, passa al setaccio milioni di siti di lavoro, analizzando quelli che chiama i "dati social" di ogni persona. Il modello di Gild cerca di quantificare e qualificare il "capitale sociale" di ogni lavoratore: calcola quindi quanto un candidato è influente nel mondo del codice e della programmazione, su siti come *GitHub* o *Stack Overflow*. Il ragionamento tende a tagliare fuori tutti i lavoratori che hanno altri impegni *offline*, che anche l'algoritmo più sofisticato non potrebbe, ad oggi, dedurre (cfr. O'Neil 2016). Se come è dimostrato dalla piattaforma ISTAT, in coppie etero formate da partner lavoratori di 25-64 anni, l'indice di asimmetria nel lavoro familiare[4] risulta essere all'incirca per il 72% sbilanciato verso le donne[5], capiamo come queste ricerche di personale possano essere *biased*: ciò implica che le donne hanno tendenzialmente più tempo impiegato nel lavoro familiare,rispetto agli uomini e per questo saranno svantaggiate dall'algoritmo di Gild.

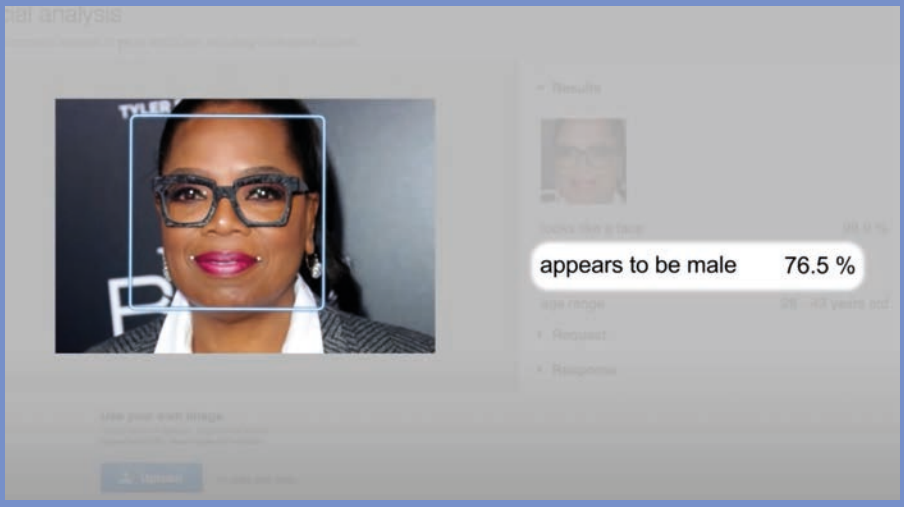
> Quando considero i modi egoistici con cui le aziende spesso usano i dati, mi viene in mente la frenologia, una pseudoscienza che fu brevemente popolare nel XIX secolo. I frenologi facevano scorrere le dita sul cranio del paziente, cercando protuberanze e rientranze. Pensavano che ogni

4 Il "lavoro familiare" include nella strutturazione dei dati Istat sia il lavoro di cura di bambini e adulti, sia il lavoro domestico, che comprende cucinare, lavare e stirare, pulizia della casa e acquisto di beni e servizi. Link: <dati.istat.it>.

5 Dati estratti il 26 agosto 2021, 16h03 UTC (GMT) da I.Stat, Link: <dati.istat.it/index.aspx?lang=it>.

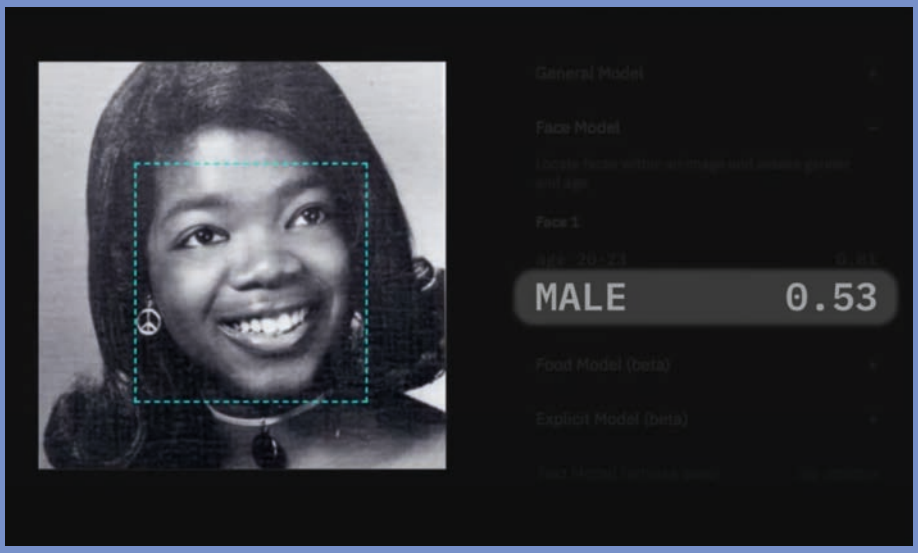


1



2

3



4

FIG. 19 Screen del video Youtube: AI, Ain't I A Woman, Joy Buolamwini Youtube, 28/06/2018.
1. Il software Face +++ non riconosce Serena Williams.
2. Il software di Amazon face recognition non riconosce Oprah Winfrey.
3. Il software Microsoft Azure non riconosce Michelle Obama.
4. IBM Watson non riconosce Oprah Winfrey da giovane.

> singola deformità fosse collegata ai tratti della personalità. Se un paziente era morbosamente ansioso o soffriva di alcolismo, la sonda cranica di solito trovava protuberanze e avvallamenti correlati - che, in cambio, rinforzava la fede nella scienza frenologica. La frenologia era un modello che si basava su sciocchezze pseudoscientifiche per fare previsioni autorevoli, e per decenni venne dato per vero, senza alcuna prova testata. I big data possono cadere nella stessa trappola. (O'Neil 2016; traduzione mia)

2.1.2 *Gender data gap*

Altro motivo per cui i sistemi algoritmici risultano distorti è l'ormai noto *gender data gap*[6], raccontato nel dettaglio da *Invisible Women. Exposing Data Bias in a World Designed for Men* (Invisibili, Einaudi) scritto da Caroline Criado Perez, evidenza come i dati di genere siano assenti, non raccolti o non presi in considerazione. I dati "invisibili" contagiano ogni settore: dall'arte alla medicina, dalla progettazione degli spazi pubblici alla prevenzione della violenza sulle donne. L'attenzione che l'autrice dedica all'assenza delle donne rappresenta il *fil rouge* delle battaglie e dell'interesse di Perez, anche in quanto giornalista. L'autrice ha raccolto diversi dati per dimostrare che viviamo in un mondo progettato per un essere umano di default: il maschio. Questo porta all'inevitabile gap a discapito di dati sul mondo femminile.

6 Con *gender data gap* si intende il divario esistente sui dati di genere. Il termine si riferisce all'assenza di dati riguardanti il genere femminile.

Il filosofo Pierre Bourdieu, in "Il dominio maschile"[7], esamina lo stato di subordinazione nella quale è sempre vissuta ed ancora vive la donna nei confronti dell'uomo. La subordinazione caratterizza la donna nell'ambito del quotidiano e le considerazioni dell'autore sono d'aiuto per comprendere meglio il gender data gap. «La forza dell'ordine maschile si misura dal fatto che non deve giustificarsi: la visione androcentrica si impone in quanto neutra», scrive Bourdieu, annettendo una nota a piè di pagina dove afferma che: «sia nella percezione sociale sia nella lingua il genere maschile appare come non contrassegnato, neutro, in qualche modo, per opposizione al femminile, esplicitamente caratterizzato» (Bourdieu 1998). Un modo per dire che il maschile è l'impostazione cognitiva di default, mentre il femminile risalta proprio per non esserlo.

7 Bordieu, partendo da uno studio antropologico sulla società androcentrica dei cabili in Algeria, dimostra la costanza della visione fallocratica del mondo nell'inconscio di uomini e donne. Versione attuale ristampata dalla Feltrinelli nell'Universale Economica.

FIG. 20 *The Library of Missing Datasets* (2016).

> L'ordine sociale funziona come un'immensa macchina simbolica tendente a ratificare il dominio maschile sul quale esso si fonda: è la divisione sessuale del lavoro, [...] è la struttura dello spazio, con l'opposizione tra il luogo d'assemblea o di mercato, riservato agli uomini, e la casa, riservata alle donne [...], è la struttura del tempo [...], con i momenti di rottura, maschili, e i lunghi periodi di gestazione, femminili. Il mondo sociale costruisce il corpo come realtà sessuata e come depositario di principi di visione e di divisione sessuanti [...]: è attraverso di esso che si costruisce la differenza tra i sessi biologici, conformemente ai principi di una visione mitica del mondo radicata nel rapporto arbitrario di dominio degli uomini sulle donne, anch'esso inscritto, con la divisione del lavoro, nella realtà dell'ordine sociale. La differenza biologica tra i sessi [...] e, in modo particolare, la differenza anatomica tra gli organi sessuali può così apparire come la giustificazione naturale della differenza socialmente costruita tra i generi e in modo specifico della divisione sessuale del lavoro. (*ibidem*)

«Gli studi di genere della seconda metà del Novecento hanno evidenziato come il modello illuminista e la sua impostazione giuridica siano costruiti su una persona identificata con un uomo, bianco, cattolico» (D'Amico 2020 - cita Facchi 2013). Il gender data gap penalizza le donne a livello professionale e nella vita pubblica, ma può costare anche in termini di salute ed è in alcuni casi questione di vita o di morte, come dimostra nella sezione dedicata alla medicina:

> Partiamo dal campo medico: da un'analisi condotta nel 2008 su una serie di libri di testo consigliati dalle «piú prestigiose

> università europee, statunitensi e canadesi» (Medical Textbooks Use White, Heterosexual Men as a «Universal Model», in «ScienceDaily», 17 ottobre 2008) risultava che, su un totale di 16.329 illustrazioni, le «parti del corpo neutre» raffigurate con immagini maschili erano tre volte piú numerose delle raffigurazioni femminili. Un'analisi di Curr-MIT , mostra invece come nel database di documentazione e gestione dei programmi universitari, è risultato che soltanto nove delle novantacinque facoltà di Medicina che immettevano dati nel sistema avevano un corso dedicato alla salute femminile. Nel 2016 uno studio sulla presenza femminile nelle ricerche statunitensi sull'Hiv (A Systematic Review of the Inclusion (or Exclusion) of Women in Hiv Research: From Clinical Studies of Antiretrovirals and Vaccines to Cure Strategies) ha reso noto che le donne rappresentavano solo il 19,2 per cento dei soggetti partecipanti alle sperimentazioni dei farmaci anti-retrovirali; mentre negli studi sui vaccini la presenza femminile sale debolmente al 38,1. (Pochettino 2020)

Pochettino prosegue, nel suo articolo intitolato "Perché l'Intelligenza Artificiale è (in)consapevolmente sessista", ricordandoci che i libri scolastici riportano le gesta di soldati, condottieri, re, politici e scienziati nella quasi totalità maschile, senza analizzare a fondo «le ragioni culturali e sociali per cui le gesta femminili sono così marginali nella cronaca dei secoli scorsi. Così nell'immaginario collettivo delle generazioni che crescono, gli scienziati sono uomini, così come la maggior parte di tutti i ruoli di responsabilità e potere. Gli algoritmi imparano. E riproducono» (*ibidem*). Un bias nel campo dei dati porta a percezioni errate, perché i dati a disposizione non sono rappresentativi della popolazione o del fenomeno di studi.

Come spiega Krishnamurthy in *Understanding Data Bias* accade anche che i dati non includano variabili che rappresentano in modo appropriato il fenomeno che si cerca di prevedere o ancora che i dati includano contenuti prodotti da esseri umani potenzialmente biased contro alcuni gruppi di persone. Secondo Criado Perez le conseguenze di questo bias si estendono in ogni ambito, dal momento che i dati sulle differenze non esistono o, se esistono, non sono presi in considerazione.

La mancanza di dati sul 50% della popolazione è presentata nel testo di Perez come una vera e propria causa della cattiva progettazione di toilette, fermate dell'autobus e altri spazi pubblici. Il testo fa riflettere sulla lavorazione dei dati, sull'importanza di conoscere chi li ha raccolti, il fine con cui sono stati raccolti ecc. Il gender gap globale di dati persiste e la nostra conoscenza del genere femminile rimane insufficiente per affrontare la sfida di progettare politiche per raggiungere gli obiettivi di sviluppo sostenibile (SDGs).

Data2X è un'associazione in prima linea nel trasformare i big data in conoscenze utilizzabili per la vita delle persone di genere femminile, indagando il potenziale di nuove e diverse fonti di dati. Il report "Big Data, Big Impacts?" (2019) riassume i risultati dei loro sussidi a supporto di attori sociali che fanno uso dei dati per colmare il divario dei dati di genere. Data2X afferma che i big data potrebbero sia fornire informazioni scandite nello spazio e nel tempo, sia offrire *insights* negli aspetti umani difficili da quantificare. Ma il potenziale dei big data sarà realizzato solo se verranno fatti investimenti complementari nei metodi di ricerca per identificare e correggere i bias.

Apart from the "language" field of application already mentioned, A.I. is currently finding practical expression above all in the field of image analysis and image creation. A subject that is traditionally particularly close to designers. (Kabel, Schultz, e Sager 2021)

2.1.3 Il design della comunicazione per la decodifica di senso

Il designer è partecipe alla diffusione degli stereotipi di genere «quando pensa, quando progetta, quando si dota di strumenti, quando manipola codici espressivi, quando produce narrazioni e mette in figura, quando produce narrazioni intorno a ciò che preesiste (che è stato progettato in passato) ecc.» (Bucchetti 2019).

Può quindi il designer intervenire per agire contro la diffusione degli stereotipi di genere? In questo senso possono gli artefatti di comunicazione diventare strumenti utili alla produzione di un mutamento e avanzamento sociale?

Il progetto della comunicazione, come afferma Zingale, «ha bisogno di scrutare il senso oltre la visione limitata e pigra dello stereotipo. Altrimenti, *ciò che si dice*, anche attraverso le forme espressive più innovative, corrisponderà sempre a *ciò che si è detto*» (Baule e Bucchetti 2013: 116).

La comunicazione visiva ha una responsabilità sociale (cfr. Frascara 2005), che riguarda l'impatto che ha all'interno del contesto sociale, il modo in cui i contenuti influenzano il pubblico, il modo di anteporre la sicurezza delle comunità all'estetica del risultato. Una certa parte degli artefatti comunicativi ostacola la crescita di una società paritaria. I temi relativi all'affermazione di una cultura della parità sono strettamente connessi al mondo della rappresentazione che andiamo a costruire, con il mondo delle immagini che quotidianamente produciamo (cfr. Bucchetti 2019).

Seguendo il filo del discorso in *Rimediazioni gender sensitive* (Caratti 2015: 50), Caratti riprende Volli. Egli sostiene che le immagini fotografiche possono essere analizzate su due livelli di analisi:

> Quello plastico della forma dell'espressione, dove la figura è analizzata in relazione a categorie topologiche (distribuzione spaziale delle figure del testo), [...] e quello figurativo in cui le immagini acquistano un senso e in cui entrano in gioco certe modalità retoriche e la conoscenza visiva che il

> destinatario ha in rapporto ai suoi schemi di pensiero, pregiudizi, stereotipi. (*ibidem*)

È la stessa Risoluzione Europea a mettere in evidenza che i comportamenti attivati dalla pubblicità (si potrebbe dedurre dalle immagini in generale) non devono essere discriminatori o degradanti quando basati sul genere[8].

La tesi segue il suggerimento che propone Caratti in *Rimediazioni gender sensitive*, vuole cioè istituire un processo di decostruzione dei testi visivi per consapevolizzare sulla struttura complessa delle immagini e di quanto le stesse rispondano a codici e retoriche predefinite. «[...] Il design della comunicazione si assume il compito ulteriore di indagare gli stereotipi che diffonde, mettendo a punto strumenti di smontaggio in grado di disvelare messaggi sottesi» (Bucchetti 2015: 32).

Il nostro obiettivo, in quanto progettisti della comunicazione, è «definire le metodologie, gli strumenti qualitativi più idonei per interpretare le immagini femminili che costruiscono il nostro quotidiano, dimostrando quanto i processi di visione e rappresentazione corrispondano verità a una mirata costruzione culturale e sociale» (Caratti 2015: 50).

Risulta infatti compito delle culture di genere studiare i fenomeni di rappresentazione di genere con il fine di creare strumenti utili a combattere i processi discriminatori. Caratti suggerisce che:

> [...] promuovere un'educazione ai media per una fruizione più consapevole delle immagini femminili da loro veicolate, comporta un ragionamento non solo sui contenuti (gli argomenti da sottoporre), ma anche una riflessione su quali possano essere i linguaggi verbo-visivi più appropriati, gli strumenti da adottare e le metodologie didattiche più adeguate. (Caratti 2015: 50)

Lo stesso si può dire per le immagini contenute nei dataset di allenamento, che vengono trasmesse alle IA. La tesi agisce decostruendo il processo di creazione e diffusione di stereotipi nei sistemi di ML: vuole riflettere quindi

8 Riferimento al punto C. della Risoluzione europea, che afferma: «Considerando che la pubblicità che presenta messaggi pubblicitari e/o degradanti basati sul genere e gli stereotipi di genere sotto qualunque forma rappresentano ostacoli per una società moderna e paritaria».

sull'uso dei registri di espressione, dei modelli retorici e della messa in scena delle immagini. Gli artefatti comunicativi hanno una natura testuale e «tali testi costituiscono la sintesi di scelte progettuali (sulla base dei diversi tipologie di segni) che generano un effetto di senso complessivo sul destinatario, proponendogli una serie di valori (più o meno desiderabili), attraverso sensi secondi e narrazioni implicite» (*ivi*: 51).

Come afferma Caratti il designer si può attivare nel processo di costruzione mediatica e può agire attraverso una rimediazione, intesa anche come:

> Porre rimedio, quello del rimediare attraverso l'azione concreta di chi opera all'interno dei processi comunicativi: gli operatori e i designer della comunicazione. Questi ultimi, attraverso il loro agire progettuale, possono contribuire a creare nuove sintesi creative, nuovi formati e nuovi linguaggi multimediali, ma allo stesso tempo possono promuovere una media education (una educazione ai media). (*ivi*: 29)

Sono svariati gli esempi di designer che hanno indagato, e successivamente agito, in merito al tema dei bias nell'IA. Sebastian Schmieg ha progettato un suo proprio algoritmo su cui basa il progetto *Decisive Mirror*, che analizza gli individui che si specchiano sulla base di tratti meno convenzionali, quale il grado di "vitalità" o di "immaginarietà". Decisive mirror vuole ricordare che gli algoritmi sono tanto accurati quanto lo sono i dati alla base e quanto lo è il design dell'algoritmo stesso. Per questo le categorie di profili che vediamo nel progetto sono in gran parte arbitrarie, casuali o imprecise (cfr. Schmieg 2019).

Charlotte Webb sfida il mondo tecnologico con il progetto *Feminist Internet*, organizzazione no-profit che sostiene la parità, l'inclusione e la partecipazione online. *F'xa* (2019) è uno degli esperimenti più noti del gruppo, consistente in un chatbot che aiuta l'utente a navigare nel mondo dell'IA biased e suggerisce modi per mitigare i pregiudizi incorporati nella tecnologia quotidiana. Il lavoro si è basato sulla ricerca di Josie Young sull'etica nell'IA (in particolare sul *Feminist Chatbot Design Process*) ed ha utilizzato gli standard del *Feminist PIA* (Personal Intelligent Assistant), dove il

femminismo è inteso come la lotta per l'uguaglianza collettiva (cfr. Jochim 2021).

Caroline Sinders, come designer e ricercatrice di ML, progetta il *Feminist Data Set* (2017 - in corso), che mette in discussione ogni fase della creazione di un sistema di IA conversazionale, attraverso domande critiche puntuali, che si focalizzano dalla raccolta dei dati all'annotazione, dall'addestramento dei dati alla scelta degli algoritmi e dei modelli. L'obiettivo di Sinders è progettare un sistema libero da pregiudizi, femminista e intersezionale (cfr. *ibidem*).

The normalizing machine è un'installazione interattiva che presenta una ricerca sperimentale nel tema del ML, che mira ad identificare la "normalità sociale". Ad ogni partecipante è chiesto di indicare la foto della persona che ritiene essere "più normale". Il lavoro automatizza il processo di discriminazione sistematica, amplificato e convenientemente nascosto sotto al black box dell'algoritmo apparentemente oggettivo (cfr. Zer-Aviv 2018). Il progetto fa riflettere anche sul meccanismo con il quale le immagini dei dataset sono spesso etichettate (vedi *Amazon Turkers*).

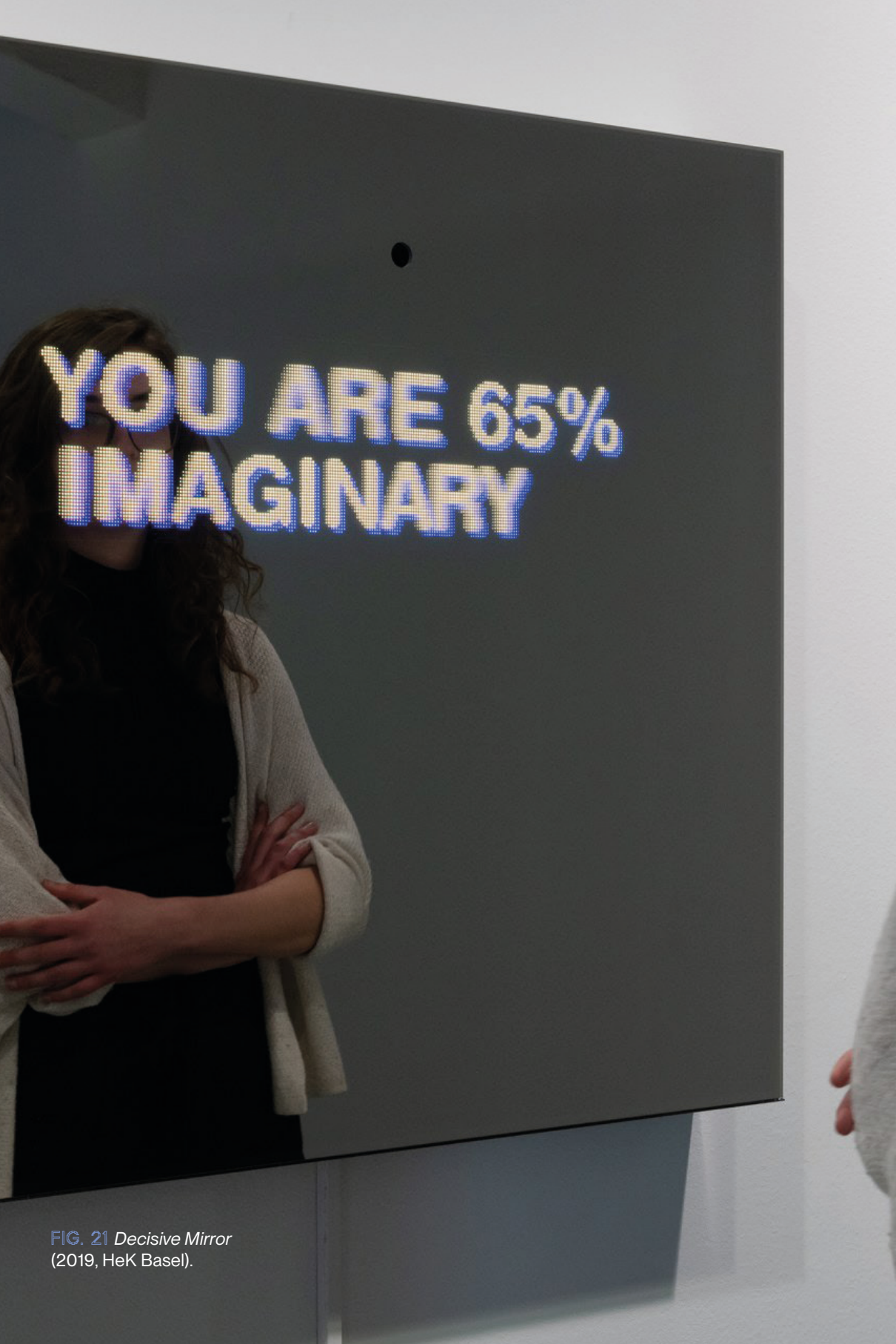


FIG. 21 *Decisive Mirror* (2019, HeK Basel).

FIG. 22 *The Normalizing Machine* (2018).

2.2 Concupiscenza: lo stereotipo e il gender bias

Maschio e femmina sono due tipi di individui che si differenziano all'interno della stessa specie (umana) ai fini della riproduzione. Possono in questo senso essere definiti solo in correlazione tra loro. Il significato della divisione della specie nei due sessi non è chiaro, infatti non si verifica universalmente in natura (de Beauvoir 2012; traduzione mia).

Se innumerevoli sono le differenze tra uomo e donna in termini biologici e strutturali, non così tante sono le differenze riportate dai dati Istat del 2018 nella percezione degli stereotipi di genere. Il report restituisce la persistenza di convenzioni radicate nella cultura popolare.

> Gli stereotipi sui ruoli di genere più comuni sono: "per l'uomo, più che per la donna, è molto importante avere successo nel lavoro" (32,5%), "gli uomini sono meno adatti a occuparsi delle faccende domestiche" (31,5%), "è l'uomo a dover provvedere alle necessità economiche della famiglia" (27,9%). Quello meno diffuso è "spetta all'uomo prendere le decisioni più importanti riguardanti la famiglia" (8,8%). Il 58,8% della popolazione (di 18-74 anni), senza particolari differenze tra uomini e donne, si ritrova in questi stereotipi, più diffusi al crescere dell'età (65,7% dei 60-74enni e 45,3% dei giovani) e tra i meno istruiti. Alla domanda sul perché alcuni uomini sono violenti con le proprie compagne/mogli, il 77,7% degli intervistati risponde perché le donne sono considerate oggetti di proprietà (84,9% donne e 70,4% uomini), il 75,5% perché fanno abuso di sostanze stupefacenti o di alcol e un altro 75% per il bisogno degli uomini di sentirsi superiori alla propria compagna/moglie. La difficoltà di alcuni uomini a gestire la rabbia è indicata dal 70,6%, con una differenza di circa 8 punti percentuali a favore delle donne rispetto agli uomini. (ISTAT 2018)

«[...] gli stereotipi sono parte del nostro sistema cognitivo, sono un atto cognitivo» (Bucchetti 2021). Lo stereotipo è inamovibile, è collocato alla radice della disuguaglianza e considerato come una «impressione fissa e immutabile,

che poco si adatta alla realtà che presume di rappresentare. Esso è il risultato della tendenza individuale a definire prima di osservare» (Villano 2003). Gli stereotipi di genere sono: «[...] immagini e rappresentazioni comuni iper semplificate della realtà che influenzano il pensiero collettivo riempiendo di specifici contenuti le convinzioni e le idee di un determinato gruppo sociale rispetto a uomini e donne e rapporti tra essi» (Ruspini 2014: 17).

La nascita della discriminazione di genere risale all'antica Grecia (cfr. D'Amico 2020), parallelamente all'idea della differenza sessuale che si concretizza tanto nella sfera naturale biologica e costitutiva di uno specifico genere, quanto nelle caratteristiche sociali e culturali. D'Amico racconta che il primo a parlare di differenza naturale dei sessi è il poeta Esiodo, in *Le opere e i giorni*, quando narra la nascita di Pandora, prima donna al mondo. «Come dice il suo nome (da pan, "tutto" e doron, "dono"), Pandora ricevette un dono da ciascuno degli dei: da Efesto un aspetto simile a quello di una 'casta Vergine'; da Afrodite la capacità di sedurre, "desiderio struggente" e "affanni che fiaccano le membra". Ma Ermes le regalò "mente sfrontata", "indole ambigua", "menzogne" e "discorsi ingannatori" (così Esiodo, Teogonia)». E continua, narrando che di fronte al:

> vaso di Epimeteo, ermeticamente chiuso [...] curiosa come tutte le donne, Pandora lo aveva scoperchiato, e dal vaso erano uscite tutte le calamità del mondo. Quando, spaventata, aveva chiuso il vaso, i mali erano già volati via, disperdendosi fra i mortali. Sul fondo era rimasta solo Elpis, la speranza. Dopo l'arrivo di Pandora, all'umanità non restava che questa. (Hesiodus 1984)

Gli stereotipi sulla discriminazione di genere originano nella civiltà greca, mentre in quella romana si «esprime in modo saldo un'organizzazione sociale patriarcale, le cui caratteristiche ritroviamo ancora oggi» (D'Amico 2020: 29).

D'Amico riprende Cantarella (2010) nell'affermare che le donne nell'impero romano erano sottoposte al potere del *pater familiae*, in forme che non garantivano nemmeno il diritto alla sopravvivenza.

> According to social role theory, gender stereotypes derive from the discrepant distribution of men and women into social roles both in the home and at work (Eagly, 1987, 1997; Koenig and Eagly, 2014). There has long been a gendered division of labor, and it has existed both in foraging societies and in more socioeconomically complex societies (Wood and Eagly, 2012). In the domestic sphere women have performed the majority of routine domestic work and played the major caretaker role. In the workplace, women have tended to be employed in people-oriented, service occupations rather than things-oriented, competitive occupations, which have traditionally been occupied by men (e.g., Lippa et al., 2014). This contrasting distribution of men and women into social roles, and the inferences it prompts about what women and men are like, give rise to gender stereotypical conceptions (Koenig e Eagly 2014). (Hentschel, Heilman, e Peus 2019)

Gli studi di Sheriffs e Mckee (1957) e di Diekman e Eagly (2000), seppur con svariati anni di distanza e a fronte di mutamenti sociali profondi, mostrano un'invarianza sui tratti che differenziano le personalità di uomini e donne[9]. Tra gli attributi associati, sia da uomini che da donne, allo stereotipo maschile troviamo: dominante, aggressivo, competitivo, indipendente, ambizioso, sicuro di sé, avventuroso e decisionista; mentre per lo stereotipo femminile: affettuosa, remissiva, emotiva, empatica, loquace, gentile (cfr. Severi 2018). Tali aggettivi vengono ricondotti ai due binomi, ampiamente utilizzati per descrivere le caratteristiche di genere, che si riferiscono a due modalità diverse dell'agire nella società: «communal and agentic» (Abele e Wojciszke 2014). Il primo, *communal*, descrive la collaborazione, la capacità di prendersi cura degli altri, la capacità di stringere relazioni interpersonali associata a tratti di calore, espressività, affiliazione (amico-nemico). Il secondo, *agentic*, rappresenta l'autonomia e l'attivismo, la tensione verso il raggiungimento degli obiettivi, la volontà di lasciare un segno associata a tratti di competenza, strumentalità, potere. Il primo termine fa riferimento a stereotipi femminili, mentre il secondo a stereotipi maschili.

La preminenza del calore sulla competenza (cioè dei tratti communal su quelli agentic) spiega l'effetto "donne meravigliose" (*wow effect*). I tratti che determinano la positi-

9 Il concetto di stereotipo è ampiamente utilizzato nell'ambito della psicologia sociale anche se il primo ad utilizzare il termine è stato il giornalista Lippmann, nel suo volume *L'opinione pubblica* (1922).

vità si riferiscono tutte a qualità di tipo communal: capacità di nutrire, aiutare, dare calore ecc. Individuano tutti tratti tipici di chi si trova in posizione subordinata, associati ad una visione tradizionale dei ruoli femminili. Ecco che l'idealizzazione del femminile che traspare dal *wow* implica un'intrinseca dichiarazione di debolezza (a dispetto della superiorità dichiarata) e di dominio maschile.

Mentre le donne sono tipicamente associate ai *communal task*, gli uomini sono tipicamente visti come *più competenti nelle cose che contano di più*. Questi stessi stereotipi hanno dimostrato di essere in gioco nella rappresentazione visiva di uomini e donne. Per esempio, in Gender Advertisements, Goffman (1979) illustra come le pubblicità ritraggono sistematicamente le donne in un modo poco serio e infantile. Uno studio di Diekman e Eagly (2000) ha individuato alcuni degli aspetti che definiscono quattro dimensioni degli stereotipi di genere: personalità positiva (ad esempio coraggioso/sensibile), aspetti cognitivi (analitico/intuitiva), tratti fisici (vigoroso/fragile), personalità negativa (arrogante/petulante).

La disuguaglianza di genere è caratterizzata e riprodotta dalla persistenza di stereotipi di genere che associano le donne a uno status sociale inferiore rispetto agli uomini (cfr. Diekman e Eagly 2000). «[...] È acclarato che vi siano nessi di causa-effetto tra stereotipi e disuguaglianze che toccano la società nella sua interezza. Perchè il rapporto, socialmente determinato, tra il fenomeno di stereotipizzazione dell'identità femminile e nel sistema di produzione, distribuzione e consumo delle immagini, attraverso i media analogici e digitali, deve essere ripensato» (Bucchetti 2021: 40).

2.3 Il gender bias nell'IA

Gli algoritmi nascono con la promessa di aggirare il problema del bias umano, portando a risultati più accurati ed equi (cfr. Kleinberg et al. 2018). Tuttavia, un crescente corpus di ricerca ha dimostrato che gli algoritmi propagano, e persino amplificano, le strutture sociali e i pregiudizi esistenti, riproducendo categorizzazioni preesistenti che si

AI News

Gizmodo: i distributori di sapone non sono in grado di identificare la pelle scura.

TOP STORIES

L'IA non sta solo imparando i nost pregiudizi: li sta amplificando.

I pregiudizi esistono nei sistemi pr gestiti da algoritmi

Perché gli uomini non credono ai d gender bias nella scienza

Una seconda ricercatrice AI affern stata licenziata da Google

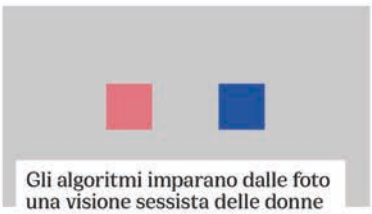
A Como stanno sperimentando il riconoscimento facciale

Riconoscimento facciale: quasi 8 p nel database di Sari sono stranieri

Riconoscimento facciale: L'Italia è nell'Ue per la tutela della privacy

FEATURED Wired.com

Siri e Cortana parlano come delle donne a causa del sessismo



Gli algoritmi imparano dalle foto una visione sessista delle donne

Quando l'intelligenza artifi vede un uomo, pensa "offic Una donna? "Smile".

WORLD NEWS

Gender bias nel machine learning per la sentiment analysis

Mike Thelwall, Statistical Cybermetrics Research Group, University of Wolverhampton, UK.

Il software di riconoscimento facciale ha un problema di gender bias

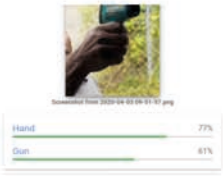
"Gli algoritmi di ricerca di immagini non solo mostrano bias nell'identificazione ma bias nel contenuto, assegnando ai politici femminili di alto livello etichette legate a uno status sociale inferiore", scrive Schwemmer.

Questioning the Fairness of Targeting Ads Online

Technology Review

Biased Algorithms Are Everywhere, and No One Seems to Care

Byron Spice, 7 luglio 2005. Carnegie Mellon University



Google apologizes after its Vision AI produced racist results

7 Aprile, 2020

1 Algoritmi difettosi stanno dando i voti a milioni di compiti di studenti.

2 Il capo della polizia di Detroit si è accorto del 96% di errori nel riconoscimento facciale.

3 Cosa succede quando un algoritmo decide di tagliare la tua assistenza sanitaria.

4 Il riconoscim facciale di A ha abbinato membri del Congresso segnaletich criminali.

Ricerca Log in

QUICK BITES

When It Comes to Gorillas, Google Photos Remains Blind

Google ha assicurato una correzione dopo che il suo software di classificazione di foto ha etichettato le persone nere come gorilla nel 2015. Più di due anni dopo, ancora il problema non è stato risolto.

ri

rocessi

lati sui

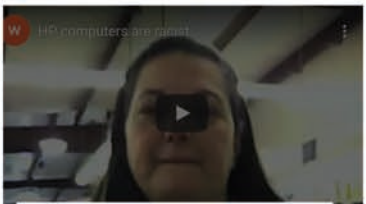
na di essere

ersone su 10

ultima

View all

ciale
ial".



WIRED: I computer HP sono razzisti

In the Outcry over the Apple Card, Bias is a Feature, Not a Bug

AI Now Institute.
22 Novembre 2019

Machines Taught by Photos Learn a Sexist View of Women

Wired, Tom Simonite. 21 Agosto 2017

nento
Amazon
28

a foto
e di

5 La Russia prova nuove tattiche di disinformazione in Africa per ampliare la sua influenza.

6 Google ha sfruttato i senzatetto neri per sviluppare l'IA di riconoscimento facciale del Pixel 4.

FIG. 23 Ipotesi facsimile di una testata giornalistica che tiene traccia di tutti gli esempi di bias nell'IA.

trovano nelle istituzioni sociali da cui l'algoritmo nasce. Ad esempio, l'elaborazione del linguaggio naturale ha dimostrato di rafforzare le associazioni di genere nel linguaggio, piuttosto che evitarle (cfr. Bolukbasi et al. 2016).

2.3.1 Osservazione

Nel capitolo precedente si sono già visti innumerevoli esempi a testimonianza della fallacità dei sistemi algoritmici nei confronti delle minoranze. Questo capitolo si focalizza su tutti quegli errori sistematici che scaturiscono in particolar modo dal gender bias. Si estendono in svariati campi: dagli annunci pubblicitari mirati, ai sistemi di assunzione, al prevedere la criminalità dal volto di una persona ecc. Di seguito alcuni esempi lampanti di come lo stereotipo di genere si insinui in modo subdolo all'interno dei sistemi di IA.

I ricercatori hanno riportato nell'articolo *Man is to Computer Programmer as Woman is to Homemaker?* la ricerca fatta tramite il modello di Word Embedding per dimostrare i pregiudizi di genere nei dati di allenamento. Il Word Embedding è un sistema che rappresenta tutte le parole attraverso vettori. Troveremo quindi che, quanto più i vettori sono vicini, tanto più le parole hanno una similarità semantica. In questo caso il dataset di allenamento è rappresentato da una serie di articoli di Google News. Le analogie sessiste restituite dal modello sono le seguenti: "Man is to computer programmer as woman is to homemaker" (L'uomo sta al programmatore di computer come la donna sta alla casalinga) e "Father is to doctor as mother is to nurse" (Il padre sta al dottore come la madre sta all'infermiera).

La ricerca si basa su due tipi di pregiudizi di genere: il bias diretto ed il bias indiretto. Nel primo le parole neutre sono, ad esempio per il femminile *homemaker, nurse, receptionist,* e per il maschile *maestro*, *skipper*, *protege*. Nel secondo, invece, le parole di genere neutro sono proiettate sull'asse della professione: ad esempio, il *softball* è considerato come una professione estremamente *per lei* e il calcio è considerato come una professione estremamente *per lui* (cfr. Bolukbasi et al. 2016).

Nel paragrafo 2.1.1.1 è stato introdotto il lavoro di ricerca Gender Shades, che illustra in modo chiaro come i sistemi di riconoscimento visivo riescano ad identificare senza difficoltà il genere di uomini bianchi, mentre falliscono nella maggior parte dei casi nell'identificare il genere di donne di colore. L'errore è da imputare alla mancanza di geo-diversità nel dataset, che parzialmente spiega perché i sistemi di computer vision etichettano una fotografia di una sposa tradizionale americana vestita in bianco come "bride", "dress", "woman", "wedding", ma una fotografia di una sposa del nord dell'India viene etichettata con "performance art" e "costume" (cfr. Zou e Schiebinger 2018).[10]

Il sistema di annunci pubblicitari mirati online è particolarmente vulnerabile a quello che è stato definito come "*algorithmic focus bias*": la ricerca *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination* scopre che «the use of Google's Ad Settings feature can lead to "seemingly discriminatory ads"[11]» (2015). Ad esempio, hanno notato che impostare il genere su femminile nella navigazione web porta ad ottenere meno annunci relativi a lavori altamente remunerativi, rispetto all'impostazione default di genere maschile (cfr. *ibidem*). Alcuni di questi risultati sono intenzionali, infatti contenuti diversi sono offerti a gruppi di persone all'interno della popolazione in base ad un particolare attributo che li caratterizza seguendo regole di marketing (cfr. Mittelstadt et al. 2019). Secondo uno studio condotto in 191 paesi del mondo, alle donne appaiono meno annunci relativi alle carriere in Scienza, Tecnologia, Ingegneria e Matematica (STEM) rispetto agli uomini (cfr. Lambrecht e Tucker 2019).

Ilinca Barsan, direttrice del dipartimento di Data Science di *Wunderman Thompson*, si è imbattuta in casualmente in un esempio lampante di bias nei dati di allenamento. Per testare il corretto funzionamento dei software di riconoscimento visivo ha caricato nell'API di Google Cloud una sua fotografia con indosso una mascherina. Il sistema ha effettivamente identificato la presenza di "Mask", con un indice di fiducia del 73,92%. Ha però rilevato anche un tag particolare: "Duct Tape", con indice di fiducia 94.51%. Barsan ha successivamente scoperto che il sistema scambiava ripetuta-

10 Di seguito vengono riportati ulteriori esempi:
1. «Facial-tracking software by Hewlett Packard did not recognise dark-coloured faces as faces» (Frucci 2009);
2. «A Nikon camera kept asking people from an Asian background: "Did someone blink?"» (Sharp 2009);
3. «An Asian man had his passport picture rejected, automatically, because "subject's eyes are closed" – but his eyes were open» (Regan 2016).

11 «L'uso della funzione Ad Settings di Google può portare ad "annunci apparentemente discriminatori"» (Amit Datta et al. 2015; traduzione mia).

mente la mascherina con il nastro adesivo. La ricercatrice ha deciso quindi di suddividere l'indagine fornendo al sistema immagini prima maschili e poi femminili, dove tutti i soggetti indossavano una mascherina. Dal totale di immagini di uomini Google ha identificato correttamente il 36% come contenente un dispositivo di protezione, mentre il 27% delle immagini raffiguranti peli facciali e un 15% raffigurante il nastro adesivo; del totale delle immagini raffiguranti donne Google ha identificato correttamente il 18% come contenente un dispositivo di protezione e un 28% raffigurante nastro adesivo, con un tasso quasi doppio rispetto a quello degli uomini.

Una ricerca di immagini di "duct tape man" e "duct tape woman" ha rivelato che, da un lato le fotografie rappresentano uomini per lo più raffigurati con il nastro adesivo su tutto il corpo mentre fanno scherzi divertenti, mentre dall'altro donne che appaiono prevalentemente in situazioni di difficoltà, con il nastro adesivo a copertura della bocca e a costrizione di braccia e mani (cfr. Barsan 2020).

I ricercatori Steed e Caliskan (2021) hanno dimostrato che, fornendo agli algoritmi di generazione di immagini, una foto di un uomo tagliata al di sotto del suo collo per il 43% delle volte si auto completerà con un corpo che indossa un abito elegante o con abbigliamento esplicitamente lavorativo. Se lo stesso processo viene ripetuto ma con una foto di una donna - sia essa anche una donna nota come la rappresentante degli Stati Uniti Alexandria Ocasio-Cortez - il 53% delle volte si auto completerà con un corpo che indossa un top scollato o un bikini.

Questi risultati hanno implicazioni preoccupanti per la generazione di immagini: simili algoritmi hanno portato ad un'esplosione di pornografia *deepfake*.

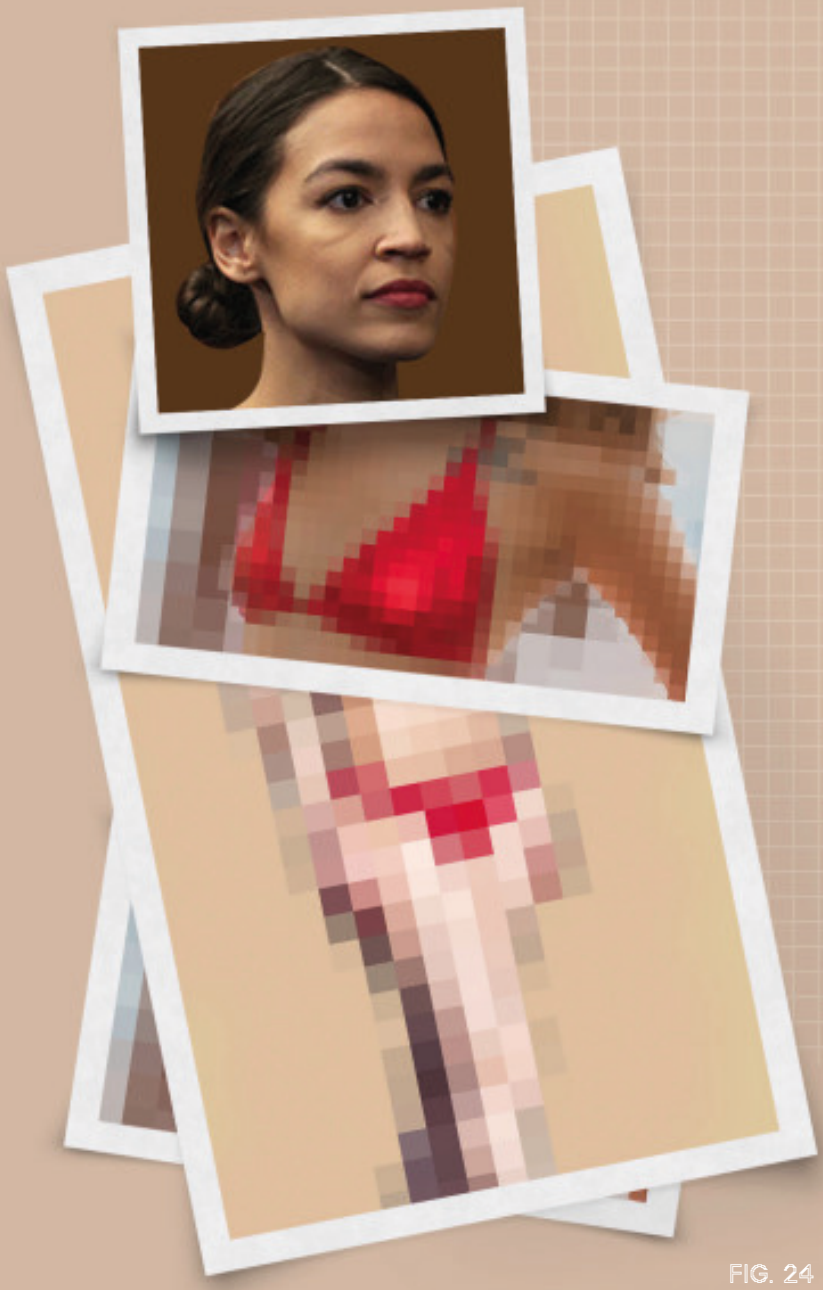


FIG. 24 Un'IA ha autocompletato una foto ritagliata di Alexandra Ocasio-Cortez con un corpo vestito solo da un bikini.

3.0 Intersezione: scienze sociali con un approccio digitale

“Spoon” is to “woman” as “tennis racket” is to “man.” At least, that’s according to AI algorithms trained on two of the more common collections of thousands of images that are usually used by researchers to help machines understand the real world.
Condliffe 2021

3.1 IA: riferimento culturale per analizzare le disuguaglianze di genere

Per inquadrare la nascita degli approcci digitali nell'ambito degli studi antropologici e sociologici risulta necessaria una brevissima sintesi della storia di Internet[1].

La prima fase, inquadrata tra il 1994 e il 2000, individua un *web* che si identifica nello *cyberspace*, ossia «a medium of pirates, pornographers, conspiracy theorists and self-publishers» (Jodi Dean *Aliens in America: Conspiracy Cultures from Outerspace to Cyberspace* 1998).

Dal 2000 al 2007 il web entra nella seconda fase: corrisponde qui ad una *virtual society*, dove «the virtual is so different and so transformative that it stands in polar opposition to the real» (Steve Woolgar *Virtual Society? Technology, Cyberbole, Reality* 2002).

Dal 2007 ad oggi: il web si identifica sempre di più con i *data*, ossia «a space for more than the study of online culture. The web as a site for the study of social and cultural phenomena» (Richard Rogers *Digital methods* 2013; Richard Rogers *Doing digital methods* 2019).

L'uso massiccio di tecnologie dell'informazione e della comunicazione rende inevitabile che sempre più attività private vengano eseguite online di cui, ad esempio, la comunicazione con altre persone o con imprese, l'acquisto di beni e servizi, la gestione di finanze, la partecipazione ad attività politiche, la richiesta di servizi alle autorità pubbliche ecc. Allo stesso tempo, la digitalizzazione di un numero in costante crescita di processi crea enormi quantità di dati o "impronte digitali", che possono essere utilizzate per fini statistici. Internet è una fonte certa di dati riguardanti la società, poichè tutto lascia una traccia (cfr. Colombo 2019).

Rogers teorizza il metodo per fare ricerca online. Per cercare le tracce e analizzare come da queste si possano ricavare informazioni statistiche è necessario, secondo

1 Sintesi tratta dalle slides di Gabriele Colombo, del corso *Integrated Course Final Synthesis Studio*, di DensityDesign Lab (2019) presso il Politecnico di Milano, Laurea magistrale in Design della Comunicazione.

Rogers, utilizzare i *digital methods*. «Digital methods are techniques for the study of societal change and cultural conditions with online data» (Rogers 2019).

Esiste una differenza sostanziale tra *virtual methods* e *digital methods*: i primi fanno riferimento ai mezzi tradizionali digitalizzati (ad esempio le *survey online*); i secondi si riferiscono ai "methods of the medium", ossia i metodi propri del mezzo di comunicazione. «Digital methods make use of available digital objects and seek to learn from how the objects are treated by the methods built into the dominant devices online» (*ibidem*). Tra noi e il passaggio dell'informazione esiste un filtro che appartiene ai devices dominanti. Quelli che Rogers chiama *digital objects* non sono altro che i link, gli hashtag, i like, gli share, i tweet, i retweet e tutti gli strumenti propri del medium. «We look at Google results and see society, instead of Google» (Rogers 2013).

Il design della tecnologia spesso cattura e riproduce concezioni di genere stereotipiche, poi ripetutamente rinforzate dalla loro reiterazione. Il modo in cui la tecnologia viene impiegata e come i dati sono raccolti ed utilizzati ha un impatto diverso sui gruppi sociali. Infatti, gli algoritmi sono *distorti* nel maggiore dei casi secondo l'etnia e il genere (cfr. Collet e Dillon 2019).

«L'intelligenza artificiale presenta, quindi, un potenziale e rilevante problema di disuguaglianza di genere» (D'Amico 2020). D'Ignazio e Klein (2020) citano l'esperimento "Word embeddings quantify 100 years of gender and ethnic stereotypes", dove computer scientist e ricercatori di storia hanno implementato il sistema di ML word embeddings per studiare l'uso di stereotipi di genere e di etnia durante il 20esimo secolo. Utilizzando dataset derivati da *Google Books* e dal *New York Times*, il team ha scovato che termini quali *intelligent*, *logical*, *thoughtful* sono tutti associati alla figura dell'uomo fino agli anni 60. Dal 60 gli stessi termini vennero sempre più attribuiti alle donne. Il fenomeno si spiega, secondo i ricercatori, con lo sviluppo del movimento femminista che avvenne tra gli anni 60 e 70. I dati rappresentano indicatori socio-culturali dei cambiamenti del patriarcato e del razzismo: devono essere interrogati in questo senso.

> Nel caso dei temi di genere, la dimensione critico-riflessiva si configura allora come l'allestimento di una vera e propria area di resistenza culturale: le rappresentazioni veicolate dai media, decostruite nel loro valore simbolico gettano luce diretta sui modelli di consumo, sui modelli relazionali, sulle forme sociali. (Bucchetti 2015: 31)

La posta in gioco è sempre più alta in relazione alla complessità che i sistemi basati sull'IA assumono. Yatskar descrive un futuro robot che, incerto di ciò che qualcuno sta facendo in cucina, offre a un uomo una birra e a una donna un aiuto per lavare i piatti (cfr. Simonite 2017).

> [...] quello che dobbiamo tenere presente sono il fortissimo impatto e le varie conseguenze nell'applicazione degli stessi algoritmi, che sono potenzialmente più dannosi rispetto a qualsiasi altro tipo di comunicazione. Sono noti infatti gli studi sul mondo della pubblicità e sull'impatto negativo della stessa nell'indurre e nel rafforzare stereotipi di tipo discriminatorio, soprattutto rispetto alle donne. Sono noti anche gli effetti di una comunicazione e di immagini sessiste rispetto al problema della violenza nei confronti delle donne, che in Italia è diventato una vera emergenza nazionale. (D'Amico 2020)

3.2 La ricerca per immagini

L'interpretazione automatica delle immagini è un progetto intrinsecamente sociale e politico, piuttosto che puramente tecnico (cfr. Crawford e Paglen). Al livello dell'immagine del dataset di addestramento si trovano presupposti e visioni del mondo. È fontamentale comprendere questa politica insita ai sistemi di IA, visto che i sistemi automatizzati si dichiarano essere la promessa di un futuro più equo. Falcinelli riflette che il funzionamento delle immagini:

> [...] non sia più un argomento "tecnico" - non in questo momento storico, bensì un problema culturale che coinvolge chiunque guardi una serie tv o scatti una foto da condividere sui social network; visto che oggi siamo tutti, almeno un po', produttori di figure. (Falcinelli 2020: 34)

Le immagini sono però oggetti sfuggevoli e stratificati, carichi di molteplici significati potenziali, infatti «[...] le immagini, oltre a rappresentare qualcosa, possiedono un meccanismo, sono dei dispositivi che funzionano in un certo modo» (*ibidem*). Interi sottocampi della filosofia, della storia dell'arte e della teoria dei media sono dedicati alla ricerca di tutte le sfumature dell'instabile relazione tra *immagini e significati* (cfr. Crawford e Paglen 2019). In alcune culture (si pensi ad esempio alla religione cristiana) l'immagine è l'emblema del potere e strumento di tutte le sue conquiste (cfr. Mondzain 2015). Viviamo nella società dell'immagine, dove la loro diffusione iper-estesa influenza il pensiero, è veicolo di conoscenza. È proprio per questo che riuscire a controllare quali immagini vengono diffuse è di primaria importanza. Di qui ne va la responsabilità del designer della comunicazione: la possibilità di arrivare ad una società paritaria è strettamente connessa al mondo della rappresentazione che il design costruisce e distribuisce poi nel "movimento perpetuo" di rigenerazione di immagini.

> Il fatto di appartenere a una società che cresce attraverso le immagini, che usa le immagini come modello, che attraverso di esse comunica, si racconta, si mostra, si distorce, che attraverso le immagini rappresenta se stessa, seppure nella sua parzialità, produce i propri luoghi comuni, i propri stereotipi visivi, che si riversano nelle pieghe della sensibilità di ciascuno, per generare fissità e pregiudizi, comporta essere calati in un movimento perpetuo, in cui le nuove immagini si generano strettamente in relazione con quelle che hanno formato sensibilità e archivi mentali di coloro i quali le hanno generate. (Baule e Bucchetti 2013)

Le immagini possono comunicare attraverso l'espressione, ossia la composizione e la regia, la quantità (che ha a che fare con la reiterazione dell'immagine stessa) ed infine attraverso il loro contenuto[2]. Esaminare le immagini nel loro senso comunicativo e sociologico è un atto complesso, considerato che:

> Ogni immagine incorpora un modo di vedere. Persino una fotografia. Perché le fotografie non sono, come spesso si crede, una registrazione meccanica. Anche se è impercettibilmente, ogni volta che guardiamo una fotografia av-

2 Dalle slides del corso di Valeria Bucchetti, 2019 *Design della comunicazione e culture di genere*, presso il Politecnico di Milano, Laurea magistrale in Design della Comunicazione.

> vertiamo l'atto selettivo del fotografo: la sua visione è stata selezionata tra un numero infinito di altre visioni possibili. (Berger 2018)

Bourdieu si è soffermato a lungo nello studio del ruolo dell'immagine fotografica, che considera importante perché «[...] una fotografia in effetti ci può fornire una versione oggettiva e verosimile - un'immagine "reale" di aspetti socialmente rilevanti di ciò che effettivamente si trova là fuori» (Goffman 2015: 72). Ma il realismo della fotografia, soprattutto se si parla di fotografia pubblicitaria o stock, è un vero è proprio assunto: si tratta di modelli che recitano una parte e vengono fotografati: avviene la finzione di un'azione in cui l'attore sa di essere ripreso (cfr. *ibidem*). «[...] Conferendo alla fotografia un brevetto di realismo, la società non fa che confermarsi nella certezza tautologica che un'immagine del reale conforme alla sua rappresentazione dell'obiettività è veramente obiettiva» (Bourdieu e Buonanno 2018).

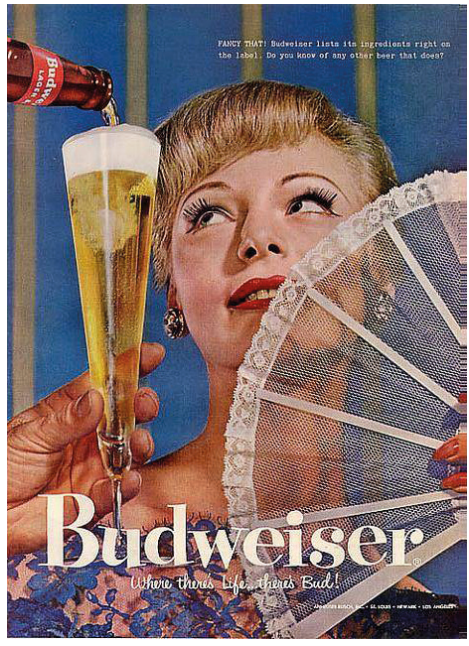
Il tema della rappresentazione di genere pone le radici negli anni 70, periodo in cui si accusavano i media di trattare l'immagine femminile in modo altamente stereotipato e conforme a figure che ostacolavano il processo di emancipazione (cfr. Caratti 2015: 33).

La rappresentazione della figura femminile nei sistemi mediali è stratificata: troviamo, infatti, immagini esplicitamente offensive, dove il corpo della donna è trattato come oggetto sessuale sottoposto al male gaze, dall'altro vi sono immagini in apparenza inoffensive. Nel secondo caso si tratta di «figure che, cristallizzando ruoli, attribuiti, riferimenti, concorrono, in forma subdolamente silenziosa, a determinare modelli quotidiani del femminile, portando il proprio contributo a sostegno dello stereotipo» (Baule e Bucchetti 2013: 29). Queste immagini corrispondono ad umiliazioni sottili, quali invisibilità e svalorizzazione, di cui non si è solitamente consci.

Il problema principale di un'analisi di genere delle immagini consiste nell'identificare il significato degli stereotipi consolidati che vengono percepiti come "normali" e quindi "invisibili". Per questo motivo, la tesi utilizza tecniche di decostruzione per analizzare i testi visivi e isolare i risultati rilevan-

1

2



3

FIG. 25 Pubblicità stampate degli anni 50 di note marche di birra e abbigliamento. 1. Schlitz "Don't worry darling, you didn't burn the beer!" 1950. 2. Van Heusen "Show Her It's a Man's World" 1951. 3. Budweiser "Fancy That" (1960), lager chiara prodotta dalla Anheuser-Busch InBev.

ti. In particolare, le funzioni estetiche, come gli aspetti grafici e visivi (colore e trattamento, elementi figurativi e tipografici) sono stati studiati in relazione alle loro funzioni semantiche per trovare i significati nascosti.

Gli strumenti di ML, come già è stato osservato, dipendono altamente dai dati utilizzati per addestrarli. In particolare gli algoritmi di riconoscimento visivi basano la loro intera "conoscenza" sui tipi di immagini che vengono loro fornite. Se le immagini sono biased, allora chiaramente lo sarà anche l'esito. Ecco che gli algoritmi addestrati con tali immagini possono mostrare la «tendenza ad associare le donne con lo shopping e gli uomini con gli spari» (Simonite 2017; traduzione mia), o ancora i database possono ritrarre donne che svolgono faccende di casa mentre gli uomini sono fuori a caccia (cfr. Zhao et al. 2017). L'IA non possiede il senso comune, la comprensione concettuale, le nozioni di causa-effetto, né tantomeno la capacità di considerare il contesto. Questo rende gli algoritmi "fragili", nel senso che non possono gestire scenari inaspettati o situazioni a loro non familiari. «Senza un ricco modello cognitivo, non ci può essere robustezza» (Guszcza et al. 2020).

Sono state condotte più ricerche a livello di dataset di allenamento sia sul testo che sulle immagini e, alcuni studiosi - soprattutto informatici e dei dati - hanno iniziato ad analizzare quello che Ferree e Hall (1990) definiscono "il primo livello di rappresentazione del bias" nelle immagini: stimare l'assenza sistematica di volti femminili in particolari arene sociali. Si veda, ad esempio, lo studio di Jia, Lansdall-Welfare e Cristianini (2015), basato sullo studio di articoli di quotidiani. I ricercatori hanno scoperto che la presenza delle donne nelle immagini varia in base all'argomento e, in particolare, le immagini politiche rappresentano soprattutto uomini. Kay, Matuszek e Munson (2015) sono giunti alla stessa conclusione nel loro studio sulle occupazioni, dove gli algoritmi dei motori di ricerca restituiscono immagini che sovra-rappresentano gli uomini rispetto al loro numero effettivo nella popolazione. Come Ferree e Hall (1990) hanno suggerito, il più basso status sociale delle donne potrebbe risultare in rappresentazioni visive di donne associate a «posizioni sociali avvilenti o marginalizzate» (*ivi*: 506).

Risulta di primaria importanza indagare le immagini che restituiamo ai modelli di ML. La ricerca non può limitarsi ad un'analisi visuale, ma deve includere uno studio semiotico, inteso come lo studio di tutto quello che può essere considerato un segno (cfr. Aiello 2020).

La pittura di Magritte è un ottimo esempio di come i segni non sono la stessa cosa di ciò che rappresentano. Lo scopo dell'analisi semiotica mira a rendere le strutture nascoste, i codici culturali sottostanti e i significati dominanti di tali testi sia visibili che intelligibili. Nel fare ciò, la semiotica è anche un potente strumento per uno studio sistematico della comunicazione visiva.

Un approccio semiotico sociale all'analisi visiva si concentra principalmente sui modi in cui determinate risorse visive possono essere impiegate all'interno e attraverso i testi per generare una gamma di potenziali significati. Le risorse semiotiche sono attivamente utilizzate sia da chi produce le immagini sia da chi le osserva: sono risorse cognitive utili a dare un senso ai messaggi visivi. Un'analisi semiotica sociale delle immagini visive deve iniziare, secondo Aiello, considerando alcune informazioni di base sul tipo di immagine da analizzare, per poi affrontare tre domande principali:

1. qual è il significato rappresentazionale di un'immagine o di un insieme di immagini? In altre parole, qual è la "storia" (o le storie) che viene rappresentata? Chi sono i "partecipanti" chiave (le persone o gli oggetti ritratti)?

2. qual è il significato interattivo di un'immagine o di un insieme di immagini? Come interagiscono le immagini con lo spettatore, per esempio attraverso lo sguardo di una persona ritratta, un certo angolo di ripresa e la dimensione dell'inquadratura?

3. qual è il significato compositivo di un'immagine o di un insieme di immagini? (cfr. *ibidem*).

Crawford e Paglen esaminano come le IA e le forme di misurazione si trasformino facilmente in giudizi morali. I

ricercatori si sono soffermati sulle tassonomie classificatorie relative all'affetto e alle emozioni umane. Lo psicologo Paul Ekman sosteneva che l'ampiezza del sentimento umano potesse essere ridotto a sei emozioni universali e, rifacendosi a queste osservazioni, i sistemi di IA stanno ora misurando le espressioni facciali delle persone per valutare tutto, dalla salute mentale, al meritare o meno l'assunzione, all'imminenza di un crimine ecc. Come sottolinea Crawford, le IA categorizzano termini quali "criminale", sottintendendo che tale comportamento risulti osservabile e traducibile in immagine.

3.3 Il femminile dal punto di vista dell'IA

Anche se ci può essere una considerevole variazione negli scopi e nelle architetture dei diversi dataset di allenamento per i sistemi di computer vision, essi condividono alcune proprietà comuni: consistono in collezioni di immagini etichettate e successivamente ordinate in categorie. Possiamo descrivere la loro architettura complessiva come generalmente costituita da tre strati: la tassonomia complessiva (l'aggregato di classi e la loro nidificazione gerarchica, se applicabile), le singole classi (le singole categorie in cui sono organizzate le immagini, ad esempio, "mela") e ogni immagine etichettata individualmente (cioè, una singola immagine che è stata etichettata come "mela"). Ogni strato dell'architettura di un dataset di allenamento è infuso di politica (cfr. Crawford e Paglen 2019).

Ogni algoritmo ha un approccio diverso per "imparare" dalle immagini del dataset, sebbene ci siano due caratteristiche principali che li suddividono: i sistemi che utilizzano il *supervised learning* e quelli che utilizzano l'*unsupervised learning* (cfr. paragrafo 1.2.2). Gli algoritmi di *computer-vision* impiegavano all'inizio il modello supervisionato, che necessita di inserire manualmente nel dataset immagini già etichettate (ad esempio le immagini di una tastiera con il tag "tastiera"). Nel 2019, la ricercatrice Kate Crawford e l'artista Trevor Paglen, hanno scoperto che queste etichette a volte contenevano un linguaggio decisamente allarmante, con etichet-

te razziste e sessiste, nonché offese insensate (ad esempio "drogato"). Si è scoperto successivamente che, anche senza l'etichettatura umana, le immagini stesse codificano modelli indesiderati: gli enormi set di dati compilati per alimentare questi algoritmi catturano tutto l'esistente in Internet, comprese la sovrarappresentazione di fotografie contenenti stereotipi spesso dannosi.

3.3.1 I confini tra scienza, storia, politica e pregiudizi nell'IA

Kate Crawford e Trevor Paglen, in "Excavating AI" analizzano i sistemi di interpretazione automatica delle immagini, affermando che ogni sistema è un progetto intrinsecamente sociale e politico, a causa di una forte asimmetria di potere (cfr. Crawford & Paglen 2019). A livello delle categorie, la concezione di genere è una semplice struttura binaria, dove "maschio" e "femmina" risultano le uniche alternative. A livello di etichetta dell'immagine c'è il problematico presupposto che l'identità di genere di qualcuno possa essere accertata attraverso una fotografia. La mostra "Training Humans"[3] esplora due questioni fondamentali: in primis come gli esseri umani sono rappresentati, interpretati e codificati attraverso il dataset di allenamento, secondariamente come i sistemi raccolgono, etichettano e utilizzano questo materiale.

Crawford e Paglen fanno un parallelismo tra l'analisi dei volti effettuata dai sistemi di interpretazione automatica e la pratica dei fisionomisti come Francis Galton e Cesare Lombroso, che studiavano le immagini di criminali, i piedi delle prostitute, misuravano i crani e compilavano archivi di immagini con misure precise, nel tentativo e nella certezza di rilevare segnali visivi nelle classificazione di etnia, criminalità e devianza dagli ideali di normalità e di borghesia.

ImageNet è uno dei più significativi dataset di allenamento nella storia dell'IA per come la conosciamo oggi. È un dataset che, come si evince dalle parole del suo creatore, ha l'ambizione di "mappare l'intero mondo degli oggetti"[4]. Il progetto nasce nel 2009 e cresce molto velocemente, fino a diventare il più grande utilizzatore del servizio di *Amazon's*

3 "Training Humans" mostra organizzata da Crawford e Paglen. La mostra ha avuto luogo all'Osservatorio di Fondazione Prada a Milano, dal 12 Settembre 2019 al 24 Febbraio 2020.

4 Fei-Fei Li, citato in Dave Gershgorn, "The Data That Transformed AI Research—and Possibly the World," Quartz, 26 Giugno 2017, link: <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>.

Mechanical Turk, con una media di 50 immagini al minuto suddivise in migliaia di categorie (cfr. *ibidem*). Alla fine, ImageNet consisteva in più di 14 milioni di immagini etichettate, organizzate in più di 20 mila categorie, tanto che per circa un decennio fu il più importante punto di riferimento nel campo di riconoscimento di immagini. In ImageNet la categoria "corpo umano" cade sotto il ramo *Oggetto naturale* > *Corpo* > *Corpo umano*. Le sue sottocategorie includono “corpo maschile”, “persona”, “corpo giovanile”, “corpo adulto” e “corpo femminile”. Qui troviamo un presupposto implicito: solo i corpi *maschili* e *femminili* sono “naturali”. C’è una categoria per il termine “Ermafrodito” che è bizzarramente situata all’interno del ramo *Persona* > *Sensualista* > *Bisessuale* accanto alla categoria “Pseudoermafrodito”. ImageNet conteneva, fino al 2019, 2.833 sottocategorie appartenenti alla categoria primaria “Person”.

All’interno si trovano categorie che mirano all’archiviazione dell’etnia, della nazionalità, della professione fino a specificare quanto buona o cattiva sia una persona, quanto sia ipocrita o tossicodipendente. Si trovano molti termini misogini e slur[5] razzisti.

> ImageNet is an object lesson, if you will, in what happens when people are categorized like objects. And this practice has only become more common in recent years, often inside the big AI companies, where there is no way for outsiders to see how images are being ordered and classified. (*ibidem*)

5 Si definisce *slur* un commento offensivo a scapito della reputazione della persona offesa.

Nel gennaio 2019, le immagini della categoria "Persona" di ImageNet hanno iniziato a scomparire. Al momento della scrittura di questa tesi, la categoria "Person" risulta accessibilie dall’interfaccia online del dataset, ma le immagini falliscono nel caricamento, quindi risultano impossibili da visionare. *ImageNet Roulette* è una delle parti dell’esibizione di Crawford e Paglen che si basa sull’interazione del visitatore che, posizionandosi davanti alla telecamera, permette all’applicazione di eseguire un rilevamento del volto. L’applicazione restituisce poi l’immagine originale con un riquadro di delimitazione e l’etichetta che il classificatore automatico ha assegnato all’immagine, attingendo alle categorie di Image Net.

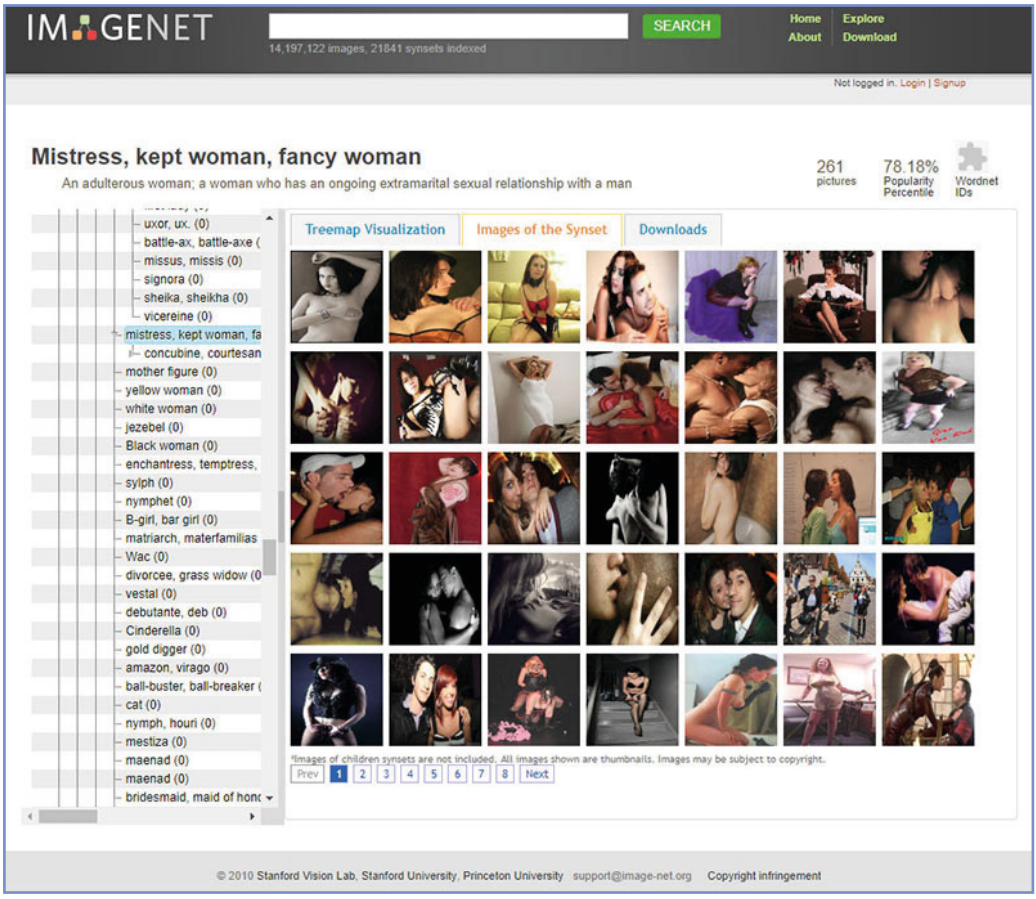


FIG. 26 Screen dalla classe “Persona” nel dataset *ImageNet*.

FIG. 27 (Pagina destra) Immagini etichettate su ImageNet (facce censurate dagli autori).


KLEPTOMANIAC


ACCUSED


GOOD PERSON


ANTI-SEMITE

3.3.2 Le immagini stock

> Imagery has become the communication medium of this generation, and that really means how people are portrayed visually is going to have more influence on how people are seen and perceived than anything else. (Jonathan Klein, 2014)

Le banche immagini diffondono e riproducono più o meno intenzionalmente modelli, molti dei quali stereotipati, escludendo gran parte delle situazioni non appartenenti alla complessità del reale[6]. Che sia a causa della loro scarsa qualità, dell'estrema insipidezza, della mancanza di veridicità, o dello sfruttamento a buon mercato, le immagini stock sono spesso ritenute insignificanti e raramente prese sul serio (cfr. Aiello 2019). Eppure, nuotiamo letteralmente in un oceano di immagini che sono state fatte per e sono distribuite da una mera manciata di corporazioni.

Giorgia Aiello studia come le immagini stock possano dare di fatto forma a specifici "modi di vedere" e come questo possa comunicarci il potere di un'immaginazione globalizzante, dove le rappresentazioni in lotti agiscono da protagoniste e si inseriscono nel nostro immaginario quotidiano. La ricercatrice afferma che le sfide chiave nella ricerca sulle immagini di stock sono legate alla loro stessa essenza, alla loro *genericità*. La gravità di questa comunicazione generica risiede nel fatto che può essere spacciata per una rappresentazione specifica e che la stilizzazione delle identità può essere mistificata nelle differenze. Il progetto "Globalization, Visual Communication, Difference", richiama l'attenzione sull'importanza culturale ed economica del lavoro di professionisti come designer, fotografi e urbanisti, sempre più responsabili della (ri)produzione di differenze culturali e sociali nella vita quotidiana.

6 Michela Rossi, *Archivi di immagini: Il contributo dei contenuti stock alla diffusione di ruoli stereotipici*, Corso di Laurea Magistrale in Design della Comunicazione, 2021, Relatrice Valeria Bucchetti, Correlatrice Francesca Casnati.

In “Taking Stock” Aiello analizza le principali banche di immagini stock quali Shutterstock e Getty Images, per esplorare abitudini e pratiche nell’uso delle immagini, ma per risalire anche alle pratiche che portano alla loro creazione. Queste immagini sono diventate la spina dorsale visiva della pubblicità, del branding, dell’editoria e del giornalismo. La nostra esposizione quotidiana aumenta in modo esponenziale in relazione all’aumento dell’uso dei social network e dei post “clickbait”[7]. Aiello, come ricercatrice interessata alla relazione tra comunicazione visiva e globalizzazione, ha svolto ricerche sulla fotografia di stock sia come industria globale che come genere visivo. Ha intervistato i fotografi che popolano le banche immagini con le loro produzioni e si è rivelato particolarmente produttivo far motivare le loro scelte fotografiche. Ad esempio, ha intervistato vari fotografi su particolari immagini da loro realizzate a cui era stata assegnata "lesbian" come una delle parole chiave (cfr. Aiello, 2013). I fotografi stock pianificano gli scatti in modo da poter ottenere diverse combinazioni di persone e concetti vendibili nel modo più efficiente ed economico possibile. Come uno dei fotografi ha raccontato nell’intervista:

> I was working with a group. Both two guys together, two girls together, then all mingling. I was trying to get as many different combinations as possible, but homosexuality did not come into my frame. There was not a relationship, not in my mind shooting it. (Aiello 2016)

Ha poi chiesto ad un fotografo di selezionare un’immagine a lui particolarmente cara tra la sua collezione di Getty Images. L’intervistato ha scelto l’immagine di un formaggio svizzero, spiegando che avrebbe voluto che quella immagine apparisse nella prima o seconda pagina nei motori di ricerca della query “cheese” ma, dato che l’algoritmo di Getty è imprevedibile, il fotografo ha dovuto continuare a pubblicare diverse versioni della stessa immagine per “truccare” il sistema e cercare di assicurarsi che almeno una di queste si posizionasse nelle prime pagine.

Le immagini stock non fanno solo parte del nostro immaginario, ma ormai sono “pasto quotidiano” delle IA, essendo i dataset in gran parte composti da immagini correda-

7 *Clickbait* è letteralmente traducibile come immagine esca, ossia un contenuto il cui scopo principale è attrarre l’attenzione e spingere i lettori a cliccare su un determinato link.

te di “watermark”, segno inequivocabile della provenienza da banche di immagini. «Therefore I hope that we keep taking stock of where this pre-produced and ready-to-use imagination leads us, both culturally and politically» (Aiello 2016).

3.3.3 I sistemi di riconoscimento visivo

I ricercatori di *Unequal Representation and Gender Stereotypes in Image Search Results for Occupations* (2020) utilizzando un dataset uniforme, composto da ritratti professionali (tratti da Wikipedia) dei membri del Congresso degli Stati Uniti, hanno interrogato il servizio di riconoscimento di immagini *Google Cloud Vision* (GCV), ampiamente utilizzato nell’industria e nella ricerca scientifica. Hanno poi replicato l’analisi su altre alternative popolari *off-the-shelf*, tra cui Microsoft Azure Computer Vision e Amazon Rekognition. In ogni piattaforma hanno riscontrato prove coerenti di due tipi distinti di pregiudizi algoritmici di genere. Gli algoritmi di ricerca di immagini non solo esibiscono *bias nell’identificazione* - gli algoritmi "vedono" uomini e donne con tassi diversi - ma anche bias nel *contenuto*: hanno assegnato ai politici femminili di alto livello etichette relative a uno status sociale inferiore (cfr. Schwemmer et al. 2020).

La figura 29 è esemplificativa dello studio fatto. Qui, la deputata del congresso Lucille Roybal-Allard è taggata con “smiling” “television presenter” “black hair,” laddove il senatore Steve Daines è etichettato con “official,” “businessperson,” e “spokesperson.” Quello che sembra essere un tag neutrale “hairstyle”, in realtà è dato a più di metà delle donne ma solo ad una ridotta percentuale di uomini. «Images of women receive about 3 times more labels categorized as “physical traits & body” (5.3 for women, 1.8 for men). Images of men receive about 1.5 times more labels categorized as “occupation” (3 for women, 4.7 for men). Images of men also receive more labels related to clothing and apparel than women» (*ibidem*).

Le etichette quali “girl” e "gentlemen" codificano in modo diretto il genere e per questo la loro corrispondenza con il genere dei membri del Congresso non è sorprenden-

te. Tuttavia, etichettare le donne adulte come "girls" mentre gli uomini con titoli più adatti all'età come "gentleman" è un vecchio tropo sessista che riemerge (cfr. Durepos, McKinlay e Taylor 2017). Inoltre, il sistema correla in modo netto il genere al bias di occupazione: nonostante tutti gli individui svolgano il medesimo lavoro, le etichette proferiscono diversamente. Di solito le donne sono etichettate con "television presenter", mentre gli uomini con varianti più autoritarie tra le quali "white collar worker," "spokesperson," e "military officer." Sebbene queste etichettature siano neutrali, «Perryman and Theiss (2013) showed that the age-diminutive "weather girl" stereotype has developed since the 1950s, when television stations began to hire non expert women as presenters to attract viewers through theatrics and sex appeal. Today, GCV labels women as "television presenter" instead of "weather girl," but the historical gender bias remains evident» (Schwemmer et al. 2020).

Goffman, inoltre, riflette sul concetto di rendere le donne tendenzialmente più giovani, associandole a bambine:

> Ne deriva quindi che nella nostra società, ogni volta che un uomo ha rapporti con una donna o con un maschio subordinato (specialmente più giovane), è abbastanza probabile che l'applicazione del complesso bambino-genitore, accorci un po' la distanza, la coercizione e l'ostilità potenziali. Ciò implica che, ritualmente parlando, le femmine equivalgano ai maschi subordinati e che entrambi somiglino al bambino. (Goffman 2015: 46)

Google, dal 2020, non consente più il rilevamento di genere nel suo servizio di riconoscimento di immagini. L'azienda afferma infatti che il genere non può essere dedotto dall'aspetto di una persona (cfr. Simonite 2020).

Kay, Matuszek e Munson citano nella loro analisi la *teoria della coltivazione*, tradizionalmente studiata nel contesto della televisione. La teoria sostiene che, sia la prevalenza, sia le caratteristiche delle rappresentazioni dei media, possono sviluppare, rinforzare o sfidare gli stereotipi degli utenti. Lo studio di Kay, Matuszek e Munson ha confrontato le percentuali di donne che sono apparse nei primi 100

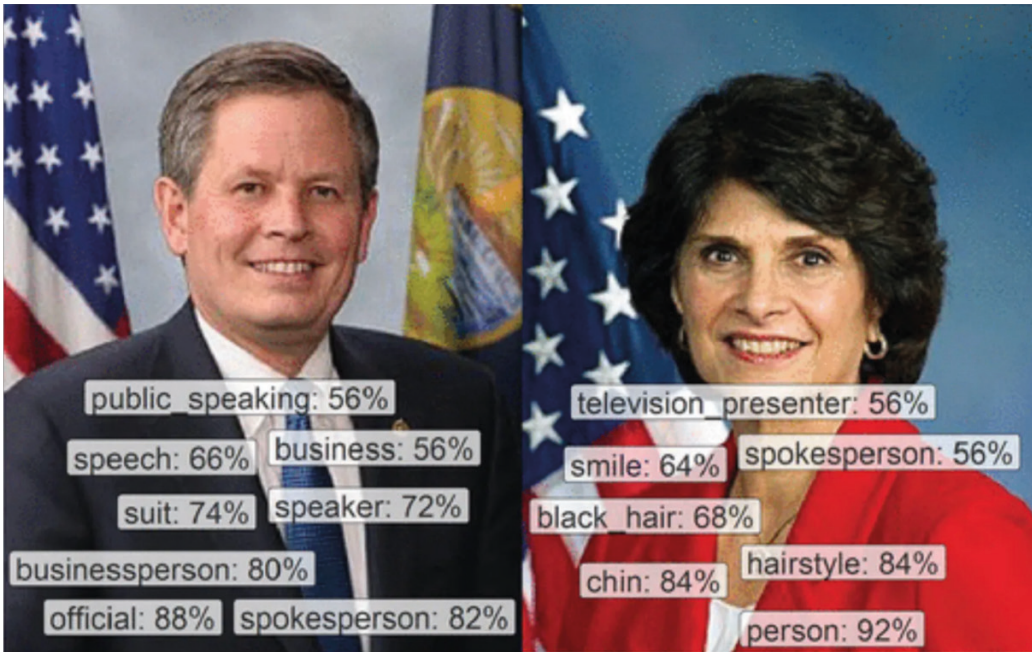
risultati di ricerca di immagini su Google nel luglio 2013 per diverse occupazioni, dal barista al chimico al saldatore, con le statistiche del *Bureau of Labor* del 2012. Il risultato trovato è stato una combinazione tra l'esagerazione dello stereotipo e la sotto-rappresentazione delle donne. In certi lavori le discrepanze erano pronunciate: ricercando CEO, l'11% delle persone rappresentate erano donne, comparate al 27% delle donne americane che nel 2012 ricoprivano la posizione di CEO. In contrasto, il 64% delle persone taggate con "telemarketers" erano donne, mentre quel lavoro è equamente distribuito tra diversi generi.

I risultati della ricerca sono distorti anche nel caso in cui le proporzioni di genere risultano rappresentative. Ad esempio, esaminando le immagini raccolte, i ricercatori hanno identificato svariati esempi di rappresentazioni sessualizzate di donne.

Tra questi compare il problema che i ricercatori hanno nominato. "sexy construction worker problem". Le immagini di donne operaie tendono ad essere caricature sessualizzate della professione. Nello studio descritto dal paper si descrive il processo di analisi: utilizzando le prime 8 immagini maschili e femminili di ogni professione, sono stati selezionati 8 aggettivi derivati dai risultati pilota e dalle domande di ricerca: *attractive*, *provocative*, *sexy*, *professional*, *competent*, *inappropriate*, *trustworthy* e *weird*. È stato poi richiesto ai turkers di indicare su una scala a 5 punti (da fortemente in disaccordo a fortemente d'accordo) se ritenessero che ogni aggettivo descrivesse la persona nella foto. Aggettivi quali *professional*, *competent* e *trustworthy* erano classificati in modo nettamente migliore quando la proporzione di persone nella BLS che corrispondeva al genere dell'immagine era più alta. Aggettivi invece quali *inappropriate* o *provocative*, avevano valutazioni significativamente più basse quando la proporzione di persone nella BLS che corrispondeva al genere dell'immagine era più alta. Gli autori chiamano questo effetto «esagerazione degli stereotipi»: le immagini che corrispondono allo stereotipo di genere di una professione tendono ad essere viste come leggermente più professionali e leggermente meno inappropriate di quelle che vanno contro lo stereotipo (cfr. Kay, Matuszek e Munson 2015).

FIG. 28 Il sistema Google image recognition tende a vedere uomini come il senatore Steve Daines come *businesspeople*, ma etichetta i legislatori donne come Lucille Roybal-Allard con termini legati al loro aspetto. Immagine di Schwemmer et al. (2020).

FIG. 29 Risultati di ricerca su Google image per "construction worker" e per "female construction worker" (Aprile 2022).


public_speaking: 56%
speech: 66%
business: 56%
suit: 74%
speaker: 72%
businessperson: 80%
official: 88%
spokesperson: 82%
television_presenter: 56%
smile: 64%
spokesperson: 56%
black_hair: 68%
hairstyle: 84%
chin: 84%
person: 92%
















4.0 Chi ha colto la mela: il gender bias nei dataset di allenamento

It is not enough to say that we need more representation in datasets. There is also a need to fundamentally question the raw data and what it reflects about society, as well as a need to design systems that can manage and fix biased data.
Collet e Dillon 2019

Gli algoritmi di ML sono potenti «costrutti socio-tecnologici» (Ananny e Crawford 2018) e in quanto tali non fanno che replicare le disuguaglianze di potere presenti nella realtà. Per questo motivo «code is both our greatest threat and our greatest promise» (Lessig e Lessig 2006).

La tesi fin'ora ha dimostrato che i bias nell'IA sono inevitabili: esistono e per contrastarli sono state trovate possibili soluzioni implementabili in livelli differenti. La prima consiste nella *sensibilizzazione*: si tratta di un problema ancora poco conosciuto, ancor meno in Italia, di cui le informazioni note sono insufficienti.

> [...] Dalla letteratura che sta ragionando sul problema dell'intelligenza artificiale e delle discriminazioni di genere risulta la consapevolezza del fatto che gli algoritmi costruiti su modelli maschili basati su stereotipi femminili rischiano di produrre conseguenze negative e che questi aspetti vanno prima conosciuti, poi denunciati e, infine, per quanto possibile, cambiati. (D'Amico 2020)

Secondariamente, le soluzioni algoritmiche odierne coinvolgono sempre più aziende e gruppi di ricerca che costruiscono software per riconoscere e valutare i rischi dei bias (si veda *AI Fairness360, Spellcheck for Bias, Ethical Bias Check* ecc.). Si tratta però di rimedi che non curano il "disagio" alla fonte, ma ne limitano solo i "sintomi". Per sradicare il problema è necessario risalire alla radice della causa e riconsiderare ogni singolo passo del *data processing*: quali domande è necessario porre, quali dati occorre raccogliere, da quale fonte, chi progetta il report, chi è incaricato di progettare il modello di ML, chi di progettare le visualizzazioni e così via. Come affermano le autrici di Data Feminism «one feminist strategy for considering context is to consider the cooking process that produces "raw" data» (D'Ignazio e F. Klein 2020).

Il punto di partenza fondamentale consiste nei dati: direttamente interrelati ai meccanismi di potere, per raccoglierli risultano necessari sia una forza intellettuale che economica non banali. Per questo motivo, è inevitabile che chi ha il potere di creare i dataset e chi di implementarli

in sistemi algoritmici influenzi i dati secondo i propri interessi e necessità. Sono quindi necessari controlli, regole, leggi per rendere i dataset quanto più inclusivi ed equi possibile. Occorrono politiche chiare, in particolar modo in un periodo storico guidato dai dati. La mancanza di uno standard è:

> [...] highlighted as a major blind spot in thinking about AI. Autonomous systems are already deployed in our most crucial social institutions, from hospitals to courtrooms. Yet there are no agreed methods to assess the sustained effects of such applications on human populations. (Crawford e Calo 2016)

Deborah Raji, collaboratrice di Mozilla nel campo della computer vision, afferma che la recente scoperta degli innumerevoli bias esistenti nel dataset dovrebbe servire come un campanello d'allarme, dato che «the actual composition of the dataset is resulting in these biases. We need accountability on how we curate these data sets and collect this information» (Hao 2021).

Il contributo della tesi consiste nell'analisi della presenza di gender bias nei dataset di allenamento e la sua successiva dimostrazione, tramite la progettazione di un foglio illustrativo, che si ispira alla letteratura attuale degli studi sull'IA. L'idea nasce da un corpus relativamente abbondante di studi in ambito di scienza dei dati (Mitchell et al. 2019, Gebru et al. 2020, Arnold et al. 2019, Kelley et al. 2009, Kelley et al. 2010, Holland et al. 2019, Bender e Friedman 2018, ABOUT ML – The Partnership on AI, Richards et al. 2020, Yang et al. 2018, Chmielinski et al. 2020 ecc.). Dal secondo decennio del ventunesimo secolo i ricercatori sopracitati propongono di utilizzare delle metodologie uniche e comuni per progettare gli standard su cui testare e presentare gli algoritmi e, prima di questi, schede standardizzate per valutare i dataset da cui i sistemi "imparano".

Nasce da *Datasheets for Datasets* ed è implementata da *Data Nutrition Label* l'idea di costruire una sorta di etichetta degli ingredienti costitutivi dei dataset, organizzata in modo rigoroso. Questa si propone di diventare una documentazione da affiancare in modo obbligatorio al dataset, utile a capirne gli aspetti fondamentali e ad implementarne

la trasparenza. Il metodo è modellato sui processi iterativi in corso per progettare gli standard di Internet (come *W3C, IETF e WHATWG*). Zou e Schiebinger (2018) sottolineano l'importanza di una scheda informativa a corredo del dataset, che spieghi come i dati sono stati raccolti e annotati e che vada ad approfondire, soprattutto nel caso di dati sensibili, le statistiche contenute all'interno:

> If data contain information about people, then summary statistics on the geography, gender, ethnicity and other demographic information should be provided. If the data labeling is done through crowdsourcing, then basic information about the crowd participants should be included, alongside the exact request or instruction that they were given. Many journals already require authors to provide similar types of information on experimental data as a prerequisite for publication. For instance, Nature asks authors to upload all microarray data to the open-access repository Gene Expression Omnibus — which in turn requires authors to submit metadata on the experimental protocol. We encourage the organizers of machine-learning conferences, such as the International Conference on Machine Learning, to request standardized metadata as an essential component of the submission and peer-review process. (Zou e Schiebinger 2018)

A livello mondiale esistono oggi varie proposte di soluzioni, che fanno riferimento a tre livelli che si articolano in diversi stratificazioni. Il report *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices* elenca in modo puntuale i livelli di riferimento:

1. il primo riguarda la *progettazione*. Le aziende dovrebbero essere proattive nell'inclusione delle minoranze nella codifica e nella progettazione delle tecnologie di apprendimento automatico. Le minoranze, tra cui quelle di genere e di etnia, devono avere un ruolo attivo nel plasmare la prossima generazione di tecnologie, in modo che gli stereotipi non vengano riprodotti e la diversità venga presa in considerazione;

2 il secondo livello è legato alla necessità di misure etiche. Andrebbero sviluppate linee guida da seguire per l'industria dell'IA, sia implementando modalità di *auditing* che verifichino i sistemi già in uso, sia cercando una maggiore trasparenza dei sistemi e dei dataset a capo di essi;

3 il terzo livello riguarda gli effetti collaterali delle strategie di digitalizzazione sul futuro del lavoro e sulle opportunità per le minoranze, quindi un'attenzione all'impatto economico politico e sociale diretto all'IA e in particolare nei lavori più precari (cfr. Morley et al. 2019).

Peters e Calvo affermano che gli sviluppatori sono frustrati per quanto poco aiuto è loro offerto dai principi ad alto contenuto di astrazione, soprattutto quando si tratta del lavoro quotidiano. Mancano delle metodologie per affrontare il problema dei bias nel mondo del codice. Gli autori riprendono Miller e Coldicott che, nella loro ricerca del 2019, osservano che il 79% dei lavoratori tecnici riporta l'esigenza di avere a disposizione risorse pratiche che li aiutino nelle considerazioni etiche (cfr. *ibidem*).

Esiste però un *gap* da colmare: gli studi che trattano o cercano di studiare in modo più o meno approfondito i dataset di allenamento dei sistemi di ML tutt'oggi non offrono metodi di analisi ed interpretazione dei gender bias nelle immagini. Si veda in particolare AI FactSheets 360 di IBM, che è uno tra sistemi ad implementare e studiare i bias nei dataset. Nel sito corporate si presenta come «uno strumento di indagine per promuovere la fiducia nell'IA aumentando la trasparenza e consentendo la governance». Nei loro esempi di etichette si fa riferimento a indicatori qualitativi generici nell'analisi di gender bias; quali la metrica di classificazione che indica quanti falsi positivi e negativi esistono. Inoltre, i dati sono indicati in modo testuale o sotto forma di tabelle senza ricorrere ad una metodologia più facilmente comprensibile quale la visualizzazione dei dati.

La tesi si basa per questo motivo sull'analisi di gender bias, con un particolare focus sui ruoli di genere, per-

ché: «[...] il processo di acquisizione dell'identità di genere sia strettamente connesso alla definizione di ruoli di genere. [...] sono i ruoli di genere ad includere comportamenti, doveri, responsabilità e aspettative connessi alla identità maschile e femminili ed essere oggetto di aspettative sociali» (Bucchetti 2015: 38).

> La diversità dei sessi è un dato di fatto ma essa non predestina ai ruoli e alle funzioni. Non esiste una psicologia femminile e una maschile impermeabili l'una all'altra, né due identità incise nel marmo. Una volta acquisito il senso della propria identità, ogni adulto ne fa ciò che vuole o ciò che può. Mettendo fine all'onnipotenza degli stereotipi sessuali, si è aperta la strada al gioco dei possibili. Ciò non significa, come ha detto qualcuno, l'instaurarsi del regno dell'unisesso. L'indifferenziazione dei ruoli non significa l'indifferenziazione delle identità. Al contrario è la condizione della loro molteplicità e della nostra libertà. (Badinter 2004)

4.1 L'anatomia di un dataset

> Datasets aren't simply raw materials to feed algorithms, but are political interventions. As such, much of the discussion around "bias" in AI systems misses the mark: there is no "neutral," "natural," or "apolitical" vantage point that training data can be built upon. There is no easy technical "fix" by shifting demographics, deleting offensive terms, or seeking equal representation by skin tone. The whole endeavor of collecting images, categorizing them, and labeling them is itself a form of politics, filled with questions about who gets to decide what images mean and what kinds of social and political work those representations perform. (Crawford e Paglen 2019)

I dati nascono come un insieme di informazioni solitamente disordinato, incompleto. Se consideriamo come esempio una fotografia di un volto, essa appare ad una macchina come una serie più o meno grande di pixel, caratterizzati ognuno da un colore identificato da un valore compreso tra lo 0 e il 256. La macchina non riconosce infatti il volto rappresentato, fintanto che non legge un'etichetta associata che afferma "questa collezione di pixel rappresenta un viso".

(cfr. Appen 2021). Il pixel è considerato la più piccola unità di informazione che compone l'immagine.

> Computer vision is not just about converting a picture into pixels and then trying to make sense of what's in the picture through those pixels. You have to understand the bigger picture of how to extract information from those pixels and interpret what they represent. (*ibidem*)

Le reti neurali e il DL infatti, sono modellati per imitare il cervello umano e riconoscere pattern. Essi interpretano i dati sensoriali etichettando e classificando gli input ricevuti. Quando un'immagine è classificata viene posta in una determinata categoria. Nella figura 31 la classificazione del primo oggetto individua "sheep" (pecora). La posizione e l'ingombro sono identificati dal riquadro che circonda l'oggetto nell'immagine. L'*object detection* (o rilevamento di oggetti) rileva le istanze semantiche di oggetti di una certa classe. Nel caso dell'immagine sottostante il sistema rileva 3 pecore e le classifica con 3 box chiamati "Sheep 1", "Sheep 2", "Sheep 3". Ogni pixel appartiene ad una classe, che in questo caso sono "sheep", "grass", o "road". I pixel della stessa classe

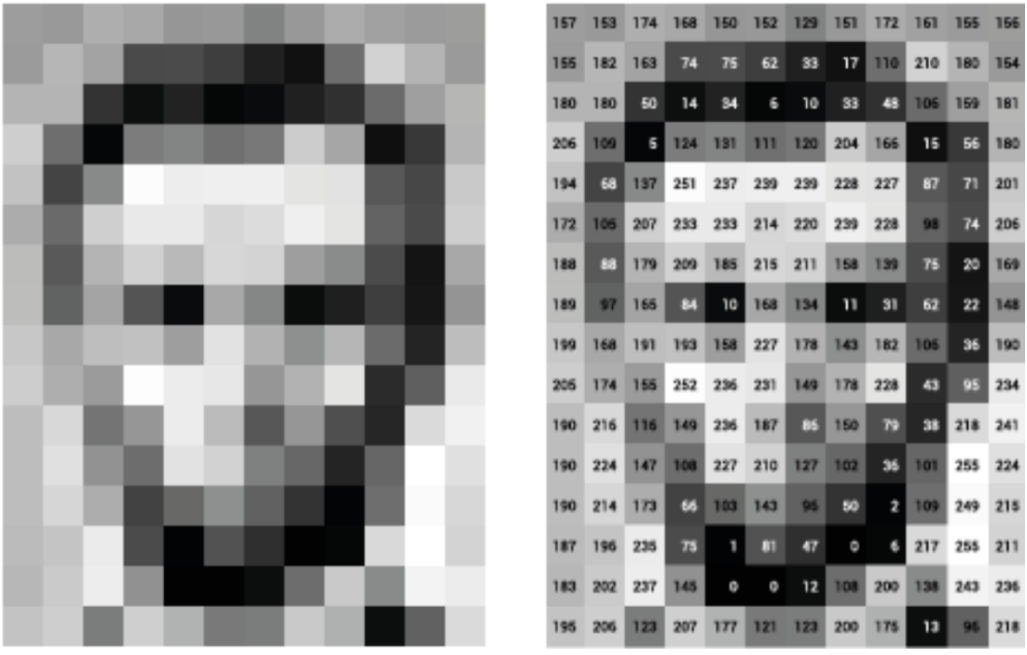


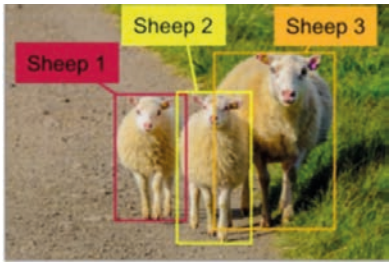
FIG. 30 Un'immagine può essere rappresentata come una matrice di valori di pixel.

sono rappresentati dallo stesso colore ("sheep" arancione, "road" grigio, "grass" verde): questa è la segmentazione semantica (cfr. Venables 2019).

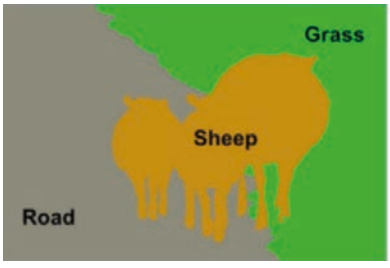
Le reti neurali e altri programmi di IA hanno bisogno di dati per rilevare i pattern. È già stato specificato che esistono diversi sistemi di allenamento dei dati, consistenti principalmente in *unsupervised* e *supervised learning*. Mentre il primo utilizza dati non etichettati, ed il modello ha il compito di trovare pattern (ossia somiglianze e differenze) nei dati per produrre inferenze e raggiungere conclusioni; con l'apprendimento supervisionato devono essere gli esseri umani a taggare, etichettare o annotare i dati secondo i loro criteri, al fine di addestrare il modello a raggiungere la conclusione desiderata (output). Questo approccio è anche chiamato "human-in-the-loop". Nel secondo caso il dataset di addestramento deve essere accuratamente etichettato prima che il modello possa elaborare e imparare da esso. Il processo è iterativo, dato che a volte l'occhio umano dell'annotatore è utile anche nel secondo step di test, dove verifica l'accuratezza delle previsioni del modello. In genere, quando si costruisce un modello, il dataset etichettato è diviso in set

FIG. 31 Differenza tra segmentazione semantica e segmentazione delle istaze.

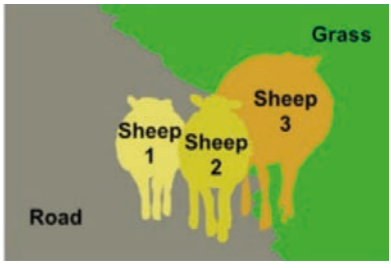
Classification + Localization



Object detection



Semantic segmentation



Instance segmentation

di allenamento e di test. L'algoritmo viene addestrato sul primo, mentre la sua performance è valutata sul secondo. Nel caso in cui il set di test non dia ottimi risultati il modello deve essere ri-addestrato.

Se ci si sofferma sul dataset di allenamento e si immagina di lavorare per una startup che deve scegliere un dataset adeguato per il proprio nuovo modello di IA, viene naturale presupporre che il punto di partenza sia una ricerca su Google, dove le parole chiave potrebbero essere: "come trovare immagini per il dataset di allenamento". In questo caso uno dei primi risultati è l'articolo *Deep Learning-Prepare Image for Dataset, A simple way to get images for your dataset*, della testata online *Towards Data Science*. Qui si afferma che le immagini per il dataset possono essere trovate tramite una ricerca su google immagini, da video di youtube, o creando il proprio personale set di fotografie.

Il messaggio comune che i vari articoli - trovati svolgendo la ricerca sopracitata - vogliono trasmettere è che risulta estremamente semplice scaricare un dataset di allenamento, ma anche costruirlo. Ed è qui che suona un campanello d'allarme: le composizioni dei dataset, infatti, non sono assolutamente regolamentate ed è altamente probabile che venga fatta una scelta errata. Si trovano online infinite liste di dataset pronti da implementare: è una giungla non organizzata e non regolamentata in cui è facile perdersi senza una bussola.

Analizzando alcuni tra i dataset di allenamento *open source*, per comprendere di quali tipi di immagine sono composti, si scopre che solo pochissimi dichiarano di avere autoprodotto le immagini, quindi di possederne il copyright. Tra questi riscontriamo il dataset *Berkeley Multimodal Human Action Database* (MHAD): contiene 11 azioni svolte da 7 soggetti maschili e da 5 soggetti femminili, nel range di 23-30 anni, fatta eccezione per un solo soggetto, più anziano. Tutti i soggetti hanno svolto 5 ripetizioni per ogni azione, in modo da ottenere circa 660 sequenze di azioni che corrispondono[1]. Altri sono i casi di dataset che utilizzano solo immagini di attori: ad esempio *TV Human Interaction Dataset* le cui immagini sono tratte da 20 diverse serie tv e consistono

1 Berkeley Multimodal Human Action Database: dataset per il riconoscimento dei movimenti umani e degli schemi di movimento. <tele-immersion.citris-uc.org/berkeley_mhad>.

in 300 clips che contengono 4 interazioni (strette di mano, dare il cinque, abbracci e baci). Un altro è il caso di IMDb Dataset, dove si trovano solo immagini di attori che compaiono nella lista *IMDB*[2].

2 IMDb Dataset: <imdb.com/interfaces/>.

La maggior parte dei dataset recupera le immagini dal "web scraping", ossia dalla ricerca di termini chiave sui più noti motori di ricerca e nella successiva implementazioni delle immagini trovate. Tra questi si trovano: Caltech 101 (Fei-Fei et al. 2004), Caltech 256 (Griffin et al. 2007), Tiny Images (Torralba et al. 2008), ImageNet (Deng et al. 2009); *ImageNet*; *COCO*; *SUN DATASET*; *CelebA dataset.* Il procedimento è quasi sempre il medesimo: vengono presi come riferimento i termini di *WordNet term* e successivamente sono scaricate le immagini relative, pescando indistintamente tra i vari motori di ricerca.

Sia ImageNet che COCO che Diversity in Images di IBM utilizzano come fonte il social media *Flickr* di *Yahoo!*. In questo modo si servono di più di un milione di fotografie senza copyright, ma coinvolgendo persone ignare.

4.2 I fogli illustrativi: casi studio

«Data practitioners evaluating datasets often have a particular use case in mind for the dataset, and are thus looking for specific issues relating to their intended use rather than browsing general information» (Chmielinski et al. 2020).

Negli ultimi anni, il campo emergente dell'etica nell'IA si è spostato verso un quadro basato su principi per giungere alla definizione, costruzione e successiva implementazione di una «trustworthy AI» (Chmielinski et al. 2020). Se si osservano i documenti che descrivono l'etica nell'IA, il principio di *trasparency* domina, seguito da *justice*, *fairness*, *non-maleficence*, *responsibility* e *privacy*. Nel report *From What to How (2020) si sottolinea* che gli sforzi compiuti fino ad oggi in termini di dati si sono concentrati sul "cosa" dell'etica, quindi dibattendo principi e codici di condotta ci-

tati, quando sarebbe più adeguato concentrarsi sul "come". I ricercatori propongono di colmare questo gap con la progettazione di una «applied ethical AI typology», di cui spiegano la metodologia di creazione all'interno del report. Sostengono, in particolare, l'esigenza di uno sforzo più coordinato da parte di ricercatori multidisciplinari, innovatori, politici, cittadini, sviluppatori e designer per creare e valutare nuovi strumenti e metodologie, al fine di garantire che ci sia un "come" per ogni "cosa" in ogni fase della pipeline di ML.

Il 22 maggio 2019, l'*Organizzazione per la Cooperazione e lo Sviluppo Economico* (OCSE) ha annunciato che i suoi trentasei paesi membri (insieme a Argentina, Brasile, Colombia, Costa Rica, Perù e Romania) hanno formalmente accettato di adottare il primo effettivo standard intergovernativo sull'IA. Lo scopo è garantire che i sistemi di IA siano robusti, sicuri ed equi (cfr. Morley et al. 2020).

Gli sforzi già citati con il fine di creare una documentazione standard per i dataset di allenamento sono più ridotti a livello istituzionale, in quanto sono soprattutto istituti privati e no-profit da cui nasce l'idea dell'etichetta citata nel paragrafo 4.0. L'etichetta consiste nella presentazione e descrizione del dataset, della sua origine e delle sue componenti. Sono svariate le proposte di metodologia di etichette e si diversificano sia in termini di contenuti (quindi i parametri scelti per le differenti valutazioni), sia in termini di User Experience. Di seguito si approfondiranno questi aspetti in diversi tipi di fogli illustrativi. Lo scopo è comprendere l'origine, il significato, le finalità, i metodi di implementazione della singola etichetta. I diversi studi sono documentati in base al loro obiettivo, alla loro metodologia, alle variabili che considerano, e all'aspetto grafico progettuale.

Caso studio 1

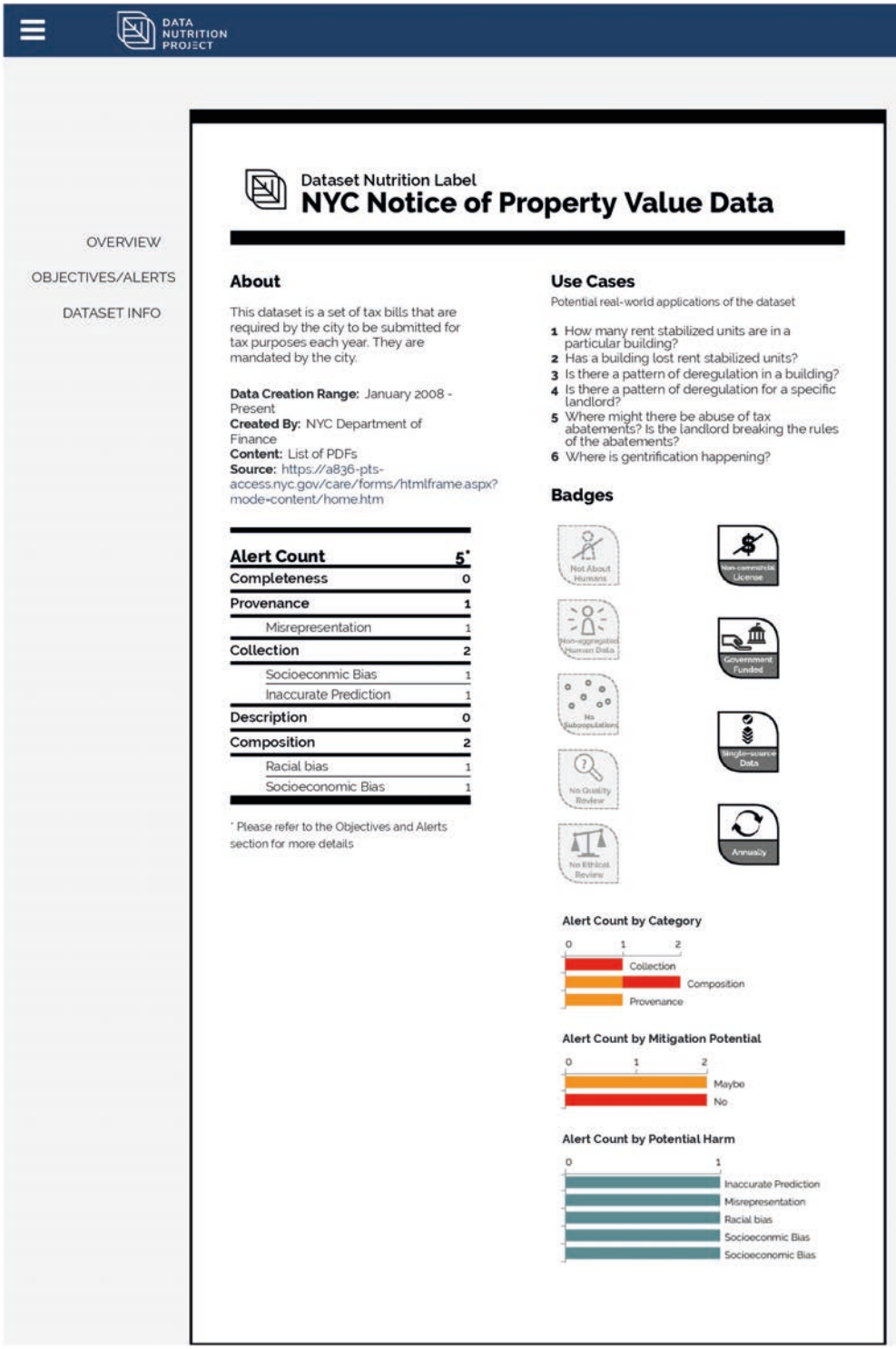
TITOLO: Dataset Nutrition Label (2nd Gen)
AUTORE: Chmielinski et al.
ANNO: 2020
FORMATO: schede interattive
TARGET: data scientist
creatori del dataset
consumatori del dataset
KEYWORD: Trasparenza, Privacy, Responsabilità, Dataset

Il progetto nasce per la necessità dello sviluppo di processi per la valutazione e l'interrogazione dei dati di allenamento delle IA. Questa è la seconda versione, che include alert di casi d'uso specifici, presentati attraverso un design aggiornato e un'interfaccia utente che ha come target finale il profilo del data scientist.

Il Dataset Nutrition Label garantisce trasparenza a livello di dataset in modo semi-automatico, con una forte enfasi innovativa su un modello di interazione che dà priorità ai problemi noti rilevanti in base al caso d'uso. Il re-design della label include:

- *Overview pane* (riquadro riassuntivo), che presenta una panoramica delle principali caratteristiche del dataset.
- *Use Cases Alerts* (avvisi sui casi d'uso), che permette all'utente di selezionare il caso d'uso e visualizzare le informazioni segnalate (Alerts e FYIs), rilevanti in modo specifico per il caso e il metodo di previsione scelto. Gli alerts presentano una scala colore a tre punti in base alla possibilità e facilità di mitigazione del problema. Gli FYIs sono codificati in verde per indicare che non c'è necessità di mitigazione.
- *Dataset Info* (panoramica qualitativa del dataset) il cui scopo è quello di fornire una documentazione per i professionisti da consultare. La panoramica è suddivisa in: descrizione, composizione, prove, raccolta, gestione.

FIG. 32 Data Nutrition Project screen.


DATA NUTRITION PROJECT
OVERVIEW
OBJECTIVES/ALERTS
DATASET INFO
Dataset Nutrition Label
NYC Notice of Property Value Data
About
This dataset is a set of tax bills that are required by the city to be submitted for tax purposes each year. They are mandated by the city.
Data Creation Range: January 2008 - Present
Created By: NYC Department of Finance
Content: List of PDFs
Source: https://a836-pts-access.nyc.gov/care/forms/htmlframe.aspx?mode=content/home.htm
Alert Count 5*
Completeness 0
Provenance 1
Misrepresentation 1
Collection 2
Socioeconmic Bias 1
Inaccurate Prediction 1
Description 0
Composition 2
Racial bias 1
Socioeconomic Bias 1
* Please refer to the Objectives and Alerts section for more details
Use Cases
Potential real-world applications of the dataset
1 How many rent stabilized units are in a particular building?
2 Has a building lost rent stabilized units?
3 Is there a pattern of deregulation in a building?
4 Is there a pattern of deregulation for a specific landlord?
5 Where might there be abuse of tax abatements? Is the landlord breaking the rules of the abatements?
6 Where is gentrification happening?
Badges
Not About Humans
Non-commercial License
Non-aggregated Human Data
Government Funded
No Subpopulations
Single-source Data
No Quality Review
Annually
No Ethical Review
Alert Count by Category
0 1 2
Collection
Composition
Provenance
Alert Count by Mitigation Potential
0 1 2
Maybe
No
Alert Count by Potential Harm
0 1
Inaccurate Prediction
Misrepresentation
Racial bias
Socioeconmic Bias
Socioeconomic Bias

Intervista: **Kasia Chmielinski**

BIOGRAPHY Kasia Chmielinski è la co-fondatrice di The *Data Nutrition Project*, un'iniziativa che mira alla costruzione di tool per migliorare l'intelligenza artificiale, attraverso dati migliori. È tecnico presso McKinsey & Company e, in precedenza, ha lavorato all'U.S. Digital Service (Executive Office of the President) e a Scratch, un progetto del MIT Media Lab (Lifelong Kindergarten Group). Ha studiato fisica all'Università di Harvard.

B.BAZZAN I wanted to know more about your role in the company - I know that you co-created the project, but beyond that what are (if you can tell) the practical tasks you do, your motivations, and your very role in the creation of the project.

K. CHMIELINSKI Thank you for reaching out. So you asked about my role in the project. We started the project in 2017 as a part of a fellowship. So my background is actually in software development, I'm a product manager. I've worked in consumer technology for about... too long. I worked at Google and I did a startup in finance. And then I worked at MIT on a kid's platform called Scratch, which is like a coding language. And then I went and I worked in the government and then I went and worked on healthcare. And so I've been building technology with data for over 10 years. Um, and so I see from the perspective of somebody who's building machines, the problem with the data becomes the problem with the machine. So for example, when I was working at Google, I lived in the UK and we were building voice search. So you could, this was in 2008, a million years ago. You could speak into your computer and you could say something and then Google would turn that into text. And it would search the internet with text. And it was pretty early days. What we saw was that we launched this thing - we thought it was really great - but it didn't work for most of the people in the UK. Because we trained it on the Queen's English, and they're not many people who speak the Queen's English, mostly it's just the queen. So the queen could do it very well, but she's an old woman. And so she didn't use it <laugh>, and everybody else did not use it. So

this was exactly the kind of situation that ended up triggering me to think, man, there's gotta be a better way to do this because the data that we use becomes the machine that we build. And then everyone is very focused on the machine and they're very focused on how the machine is sexist or racist or biased. The machine's not really anything, the machine is just a mirror of what you fed it. And so that analogy of feeding and nutrition, food, health ended up kind of getting twisted into this nutrition label, which I don't know if in Italy you have the same thing, but here on food packages, there's a nutrition label on the back and it tells you what's inside. And so that's the analogy that we created as part of this fellowship. It was myself and a few other people (they're much smarter than I am). But I was the product lead for that project. So I ended up becoming kind of the leader with the hat, making sure everybody was following me, because there were a bunch of people like researchers, artists, data scientists and engineers and all of us were doing this together, but in our free time.

So we had the idea: what would it look like to build a nutrition label for datasets? The fellowship ended and we had a prototype and a paper. And then everybody just went back to their lives, but I still really loved the project. And there were a few of us who still really loved it. And so at that time, the project was probably six people, three of them went away and then three other people from a different fellowship joined us.

That's kind of been the way that the team works is that people stay for as long as it makes sense. We don't really get paid. So the value has to be the value of the work. Like right now we have two kinds of fellows or researchers who are joining us one from policy, another one, who's a data scientist from Germany. And we just kind of have this rotating cast of people who come in and help us.

We had this version of the label and now we're working on the third version of the label. I'd say the first version, which was the prototype, was very quantitative and

it was here are the distributions of all of the variables. And here are some key things that you should know about the dataset. The second version was very qualitative (it's on the website now). It's really like 35 questions. And you ask that of the person who knows the data set the best. And it's not very quantitative at all. Some of the questions are binary or some of the questions have kind of a multi variable answer and those are set. So you have to choose from a list, but that's pretty much the only structure. We're using a lot of the questions from another brilliant group in the space called Datasheets for Datasets (Microsoft) and we've used a lot of their questions.

B.BAZZAN What, in your opinion, differentiate your label from the other ones?

K. CHMIELINSKI I think the piece on top of this that makes us maybe different from other people is that we really care about the experience of the label. So we're thinking about the design, we're thinking about how someone is actually going to use it. I think the challenge with something like datasheets is that it's a, it's a paper. It's this documentation, a PDF that comes with a dataset. And I know from my own experience when I'm on deadline and my boss says to me, you have till Tuesday I'm not gonna read the documentation. So the second version was very qualitative and the third version is going to be kind of a combination of a quantitative in the qualitative. What we have been doing for the last year has been to go and get feedback from a lot of different people. What we heard was there really two types of users: one who wants to understand the data mm-hmm and the other that wants the use, the data.

And obviously those are not mutually exclusive. But generally speaking, someone who wants to understand would be like a business owner or kind of a non-technical product manager or journalist or policy person, and someone who's gonna use the data as like a data scientist or an analyst who's really gonna dive in and build something. And that's really two different per-

spectives. And so what we've started to think through is on the understanding side, you just need a high level snapshot similar to the food, you have fat sugar carbohydrates and ect. And I think we're getting there with the current version, with the badges, those kinds of things. And then on the use side, the use will be more like: how do you actually use this data? What are the specific practical things about the variables: about the distribution and sensitive attributes, like gender. That's really kind of like known technical issues.

You mentioned gender and specifically gender bias. This is really interesting because I was on a panel a few weeks ago and someone asked me: "Hey, you come from the data nutrition project. Cool. How should we think about gender? What are the right gender categories?" And I was like, okay, that's a great question. We will never answer that question because our job is not to tell you how to do gender. It is to provide for you a framework so that you can see how that is managed in the data. What we're trying to do is provide transparency into the data set. And I think that that is a very fine line. We might be able to say some things that other people have done, but we're not gonna say the way to code gender would be to have male, female non-binary and an open text field, because in five years it's probably totally different. Same with race. For example, race is so specific to a country in the US. There's maybe one way to do it, but in Italy it might be completely different. Why should I be the one who determines that.

B.BAZZAN It's really funny, because while I was explaining this project a friend of mine asked me the very same question, and I responded in a similar way that it isn't my spot to determine how to behave?

K. CHMIELINSKI I think that's the thing though, is that everybody just wants to know the answer, and the point is that sometimes there's no right answer. There is maybe a better answer for what you're doing. So I was talking to somebody yesterday about healthcare and healthcare data sets. Yeah. And we were talking

about the gender field on a dataset about prostate cancer. You can only get prostate cancer if you have a prostate, which is a male bodied person, or somebody who is born with male genitalia would have that, but their gender could be anything. Now, whether or not you even need that is unclear, you might not need that data. So, so when someone says, oh, let's make sure that this data set is representative and has equal distribution of sex and gender. You might not need that. It really depends on what you're trying to do with the data. And if you're trying to do something with the data, then you have to think about what is the right representation or sample for what you're trying to do. It's not that every, it's not that the same data set will need to be modified in the same way for every use case for each use case. You need to think about how the data is good or bad for that.

B.BAZZAN What were the steps towards the standardization of the process, because you mentioned that, for example, the badges are really specific, but you can use them in a lot of dataset. So this step is really interesting to me. How did you manage to do it?

K. CHMIELINSKI I mean, I wish I could say it was scientific, but it was kind of a combination of looking at what others in the space have done and then drawing on our own experience of being data practitioners, scientists and analysts. I think the third component there is thinking about what other people do? What do we know is useful? What is possible to standardize? Datasets are so different. I could have a data set about trees in New York, and I could have a dataset about people in prison. And those are both datasets. So how can I make sure that the things that I have that are standard will actually be applicable for trees as well as people in prisons? So the third component is what is actually universal. And the fourth thing was probably actually one of the more helpful things, which was what matters when we think about harm. So what are the key things that you need to know, if you're going to use that dataset, to make sure that you minimize harm. And it turns out that

> although some people care about animals or plants or roads or something, most people are concerned about hurting people. Most people are concerned when it comes to AI and bias about negatively affecting communities of people. That's kind of how it is. So we said, well, when it comes to what standardized elements we want to pull out, it's probably about people. It's probably which communities are affected by consent privacy, who is funding this, where is it coming from... So anything it has to do with information that could help minimize harms or highlight it.

B.BAZZAN Really interesting process. And there's also a step of standardizing it even more because I saw that in the website you mention creating a tool that is like a fill form that would help a lot of people. It would be useful specifically for the creators of the dataset to implement this label on their website and on their data sets?

> **K. CHMIELINSKI** That's a real challenge. From my perspective of being a product manager, you're never done building software for companies. It's been an interesting process for me to be like, "Kasha, one day you have to be done". You have to be done with the standard, because then you can build the label maker (that's what we call it) and the comparison tool. And I just keep going on to the next version. So there'll be a label maker that will help you actually in generating the label, that will package up all those questions, in a very easy way like a wizard basically that walks you through it. But we can't do that until we've finished the label.

B.BAZZAN How do you think that a graphic designer could help you in this mission? Do you think that a UI/UX would be helpful for this particular project?

> **K. CHMIELINSKI** It's so important. It's incredibly important because I think that is, like I said before, the "differentiator", one of the things that we do that most others are not doing. I think about the presentation and how this will be interacted with, by data owners, creators, data scientists, journalists and policy makers and all that.

So that's actually incredibly important. And right now the lead person who's in charge of the next version of the label is a designer. (Jess Kosky) And she is actually walking through that process with us, saying "what was good in the last version? What could be better?". The role of the UX designer here is essential. Once we have the UX design ready, we're gonna need to find a UI designer. We don't have one yet, but we'll find one somewhere. And hopefully someone awesome can join us for that and help us make it look really good. Finally we have a tech team that will actually build it, obviously.

B.BAZZAN How did you test the usability of the label??

K. CHMIELINSKI We talked to over 20 different people to get the feedback on the label. We ran a few sessions where we actually gave people scenarios. So we gave them two different labels and we said: "you're going to need to choose one of these data sets to use in this way. Here are the labels. Can you make the choice of which dataset based on the label?"

B.BAZZAN Do you think it would be useful to implement data visualization on the label? Because I think it's really helpful, especially for people that don't really know about this topic. Usually the data vis helps to realize what the focus is.

K. CHMIELINSKI This is kind of the component that we're hoping to bring into the third version. It'll be kind of qualitative and quantitative. I'm not an expert in this, but it feels like to me, the most important thing to visualize is probably some of your basic descriptive stats. So how big is the data set? How much is missing? Um, probably some distributions of key variables. So if you do have sensitive attributes such as race, gender, income, disability status, whatever that you visualize those, or enable some kind of an interactive module and select from sensitive, because I think the sensitive attributes are really the thing that most people care about. And then if you really want to dig into the data, you can do that somewhere else. I don't think that we are a data exploration tool, so I don't think we need to be that

> powerful, but I think that there's some elements that we might want to bring out because visually it's easier than writing a paragraph. I think there's definitely value in data viz, for sure. Um, but I don't know how to do it. I don't know exactly which things to pull out or how to visualize it. What were you thinking in terms of what would you visualize or, you know, for what purpose?

B.BAZZAN I was thinking about the second part of the label, like the objectives and the alerts. That could be really interesting, because you have the selection part and each topic you select has specific alerts and that could be really interesting to visualize for me. So that was I was thinking.

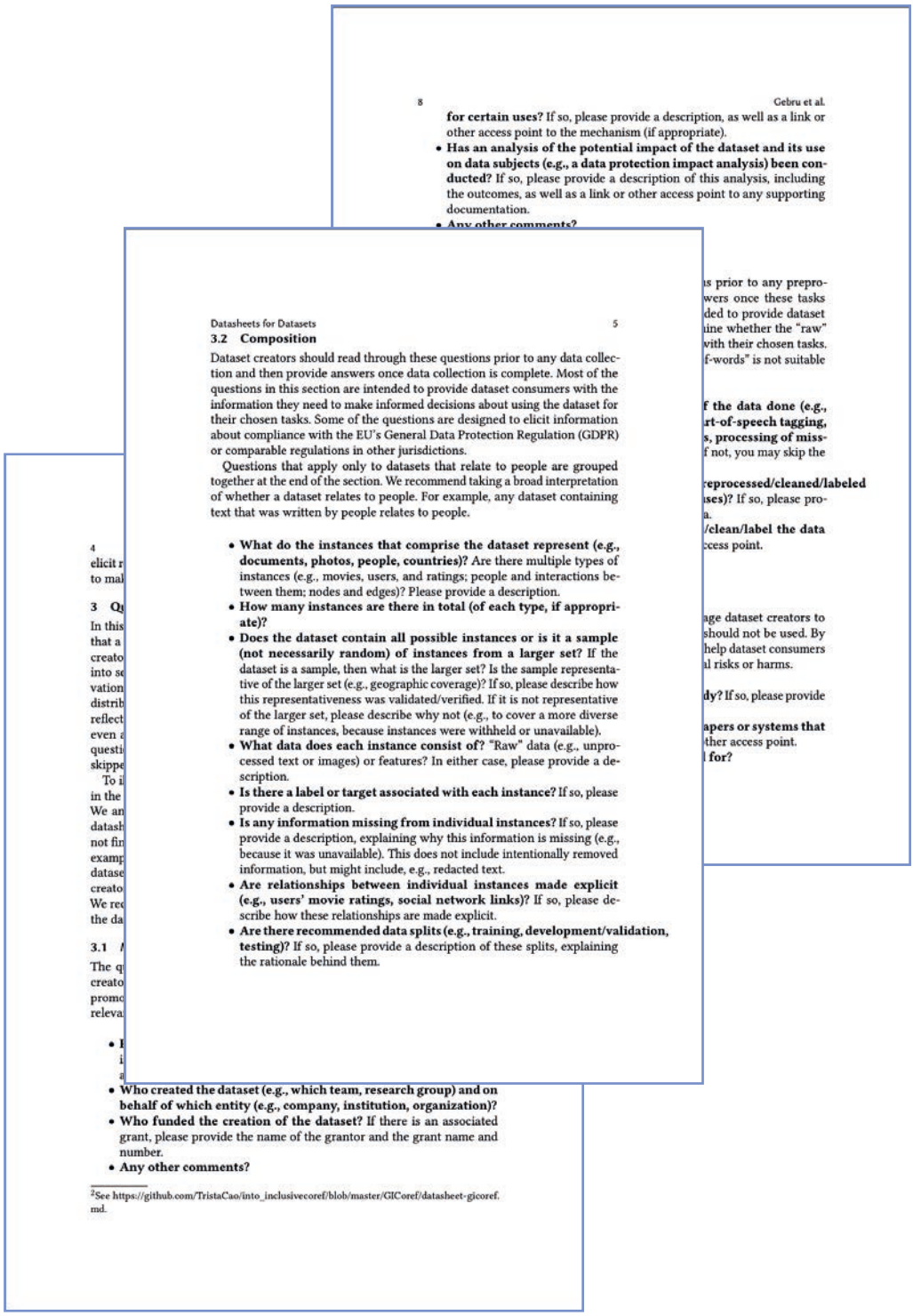
> **K. CHMIELINSKI** I think that's interesting. And we had a lot of conversations about how to visualize alerts because alerts are also many different kinds of things. You can have alerts like "you can't use this for commercial purposes", or "the representation of Asians in this dataset is very low". It's like two totally different kinds of alerts. Anyway, this is part of a longer conversation.

Caso studio 2

TITOLO: Datasheets for Datasets
AUTORE: Gebru et al.
ANNO: 2020
FORMATO: file di testo
TARGET: creatori del dataset, politici, consumatori del dataset, individui inclusi nei dati
KEYWORD: Trasparenza, Privacy, Responsabilità, Dataset

Datasheets for Dataset nasce con l'obiettivo di creare un processo standard per documentare i dataset e per rispondere alle esigenze di due stakeholder chiave: i creatori ed i consumatori dei dataset. Per i primi l'obiettivo principale è incoraggiare ad una riflessione sul processo di creazione, distribuzione e mantenimento, compresi i potenziali rischi o danni e le implicazioni d'uso. Per i secondi l'obiettivo è invece garantire che ricevano le informazioni necessarie che permetta di compiere una scelta conscia del dataset. Il metodo si ispira al foglio illustrativo di un oggetto di elettronica, dove vengono illustrate le singole componenti insieme alle sue caratteristiche operative, ai risultati di vari test di sforzo, agli usi raccomandati. Lo scopo è quindi facilitare la comunicazione tra i creatori dei dataset e i consumatori degli stessi, incoraggiando le comunità di ML a prioritizzare trasparenza e responsabilità. Le domande si suddividono in: motivation, composition, collection process, preprocessing/cleaning/labeling, uses, distribution, maintenance.

FIG. 33 Datasheets for Datasets screen.



Datasheets for Datasets 5

3.2 Composition

Dataset creators should read through these questions prior to any data collection and then provide answers once data collection is complete. Most of the questions in this section are intended to provide dataset consumers with the information they need to make informed decisions about using the dataset for their chosen tasks. Some of the questions are designed to elicit information about compliance with the EU's General Data Protection Regulation (GDPR) or comparable regulations in other jurisdictions.

Questions that apply only to datasets that relate to people are grouped together at the end of the section. We recommend taking a broad interpretation of whether a dataset relates to people. For example, any dataset containing text that was written by people relates to people.

- **What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.
- **How many instances are there in total (of each type, if appropriate)?**
- **Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).
- **What data does each instance consist of?** "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.
- **Is there a label or target associated with each instance?** If so, please provide a description.
- **Is any information missing from individual instances?** If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.
- **Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?** If so, please describe how these relationships are made explicit.
- **Are there recommended data splits (e.g., training, development/validation, testing)?** If so, please provide a description of these splits, explaining the rationale behind them.

- **Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**
- **Who funded the creation of the dataset?** If there is an associated grant, please provide the name of the grantor and the grant name and number.
- **Any other comments?**

[2]See https://github.com/TristaCao/into_inclusivecoref/blob/master/GICoref/datasheet-gicoref.md.

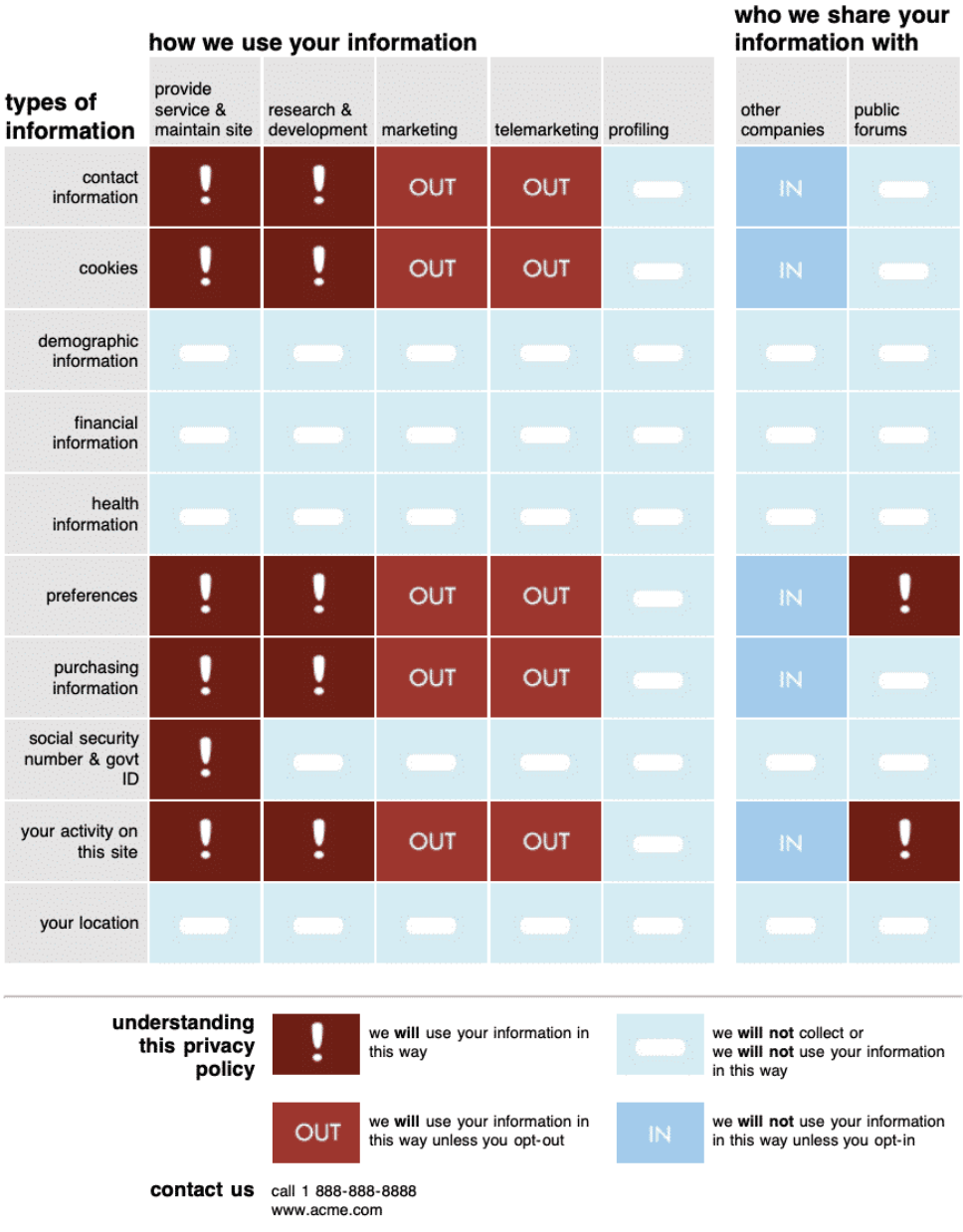
Caso studio 3

TITOLO: A “Nutrition Label” for Privacy
AUTORE: Kelley et al.
ANNO: 2009
FORMATO: infografica statica
TARGET: individui coinvolti nelle classifiche, stakeholders, sviluppatori
KEYWORD: Dati personali, Trasparenza, Privacy, Facilità di comprensione

“Nutrition Label” for Privacy si basa su un processo di design iterativo per sviluppare un’etichetta di privacy che presenta i metodi con cui le organizzazioni raccolgono, usano e condividono le informazioni personali. Questa ricerca affronta la lacuna nella comunicazione e nella comprensione delle politiche di privacy, creando un design informativo che migliora la presentazione visiva e la comprensibilità delle politiche di privacy. Le fonti di ispirazione principali sono due: da un lato si attinge alle etichette di alimenti, di avvertimenti e di consumo energetico; dall’altro allo sforzo fatto per la creazione di un’etichetta standardizzata della privacy nel contesto bancario.

Il documento A “Nutrition Label” for Privacy (Kelley et al. 2009) descrive la metodologia di progettazione, i risultati di due focus group ed infine i risultati di accuratezza, tempistica e gradevolezza derivati da uno studio di laboratorio con 24 partecipanti. I risultati dimostrano che, rispetto alle politiche di privacy esistenti, l’etichetta proposta permette ai partecipanti di trovare le informazioni più rapidamente e, allo stesso tempo, fornisce un’esperienza di ricerca delle informazioni più immediata.

FIG. 34 A “Nutrition Label” for Privacy. Immagine di Kelley et al.

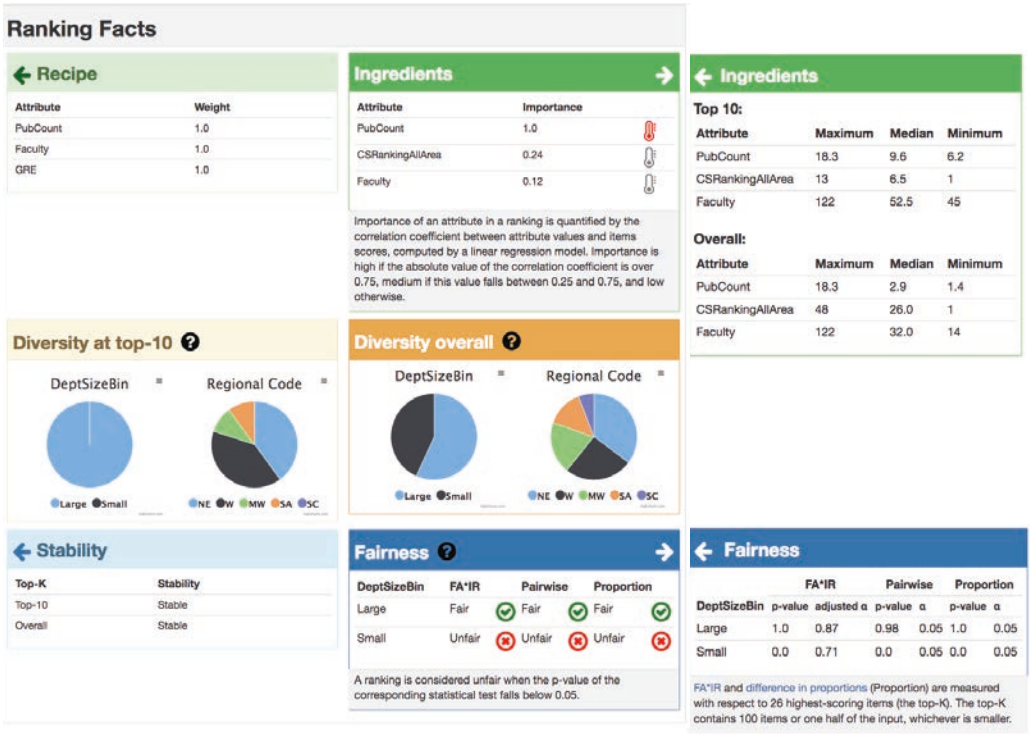

The Acme Policy
how we use your information
who we share your information with
types of information
provide service & maintain site
research & development
marketing
telemarketing
profiling
other companies
public forums
contact information
cookies
demographic information
financial information
health information
preferences
purchasing information
social security number & govt ID
your activity on this site
your location
OUT
IN
understanding this privacy policy
we will use your information in this way
we will not collect or we will not use your information in this way
we will use your information in this way unless you opt-out
we will not use your information in this way unless you opt-in
contact us
call 1 888-888-8888
www.acme.com

Caso studio 4

TITOLO: Ranking Facts
AUTORE: Yang et al.
ANNO: 2018
FORMATO: applicazione web based
TARGET: individui coinvolti nelle classifiche, stakeholders, sviluppatori
KEYWORD: Graduatorie, Trasparenza, Inclusività, Valutazione dei casi d'uso

Le decisioni algoritmiche spesso si traducono in punteggi e classifiche di individui per determinare innumerevoli concetti: dal merito di credito o dalle ammissioni al college fino alla compatibilità con partner. Oltre al problema di discriminazione, i risultati delle classifiche sono spesso instabili, dato che piccoli cambiamenti nei dati o nella metodologia di classificazione possono portare a cambiamenti drastici nell'output, rendendo il risultato facile da manipolare. Ranking Facts è costituito da una serie di widget che presentano i risultati di ricerca sui temi di fairness, stability e transparency e che comunicano i dettagli della metodologia di classificazione all'utente finale.

FIG. 35 Ranking Facts screen. Immagine di Yang et al 2018.

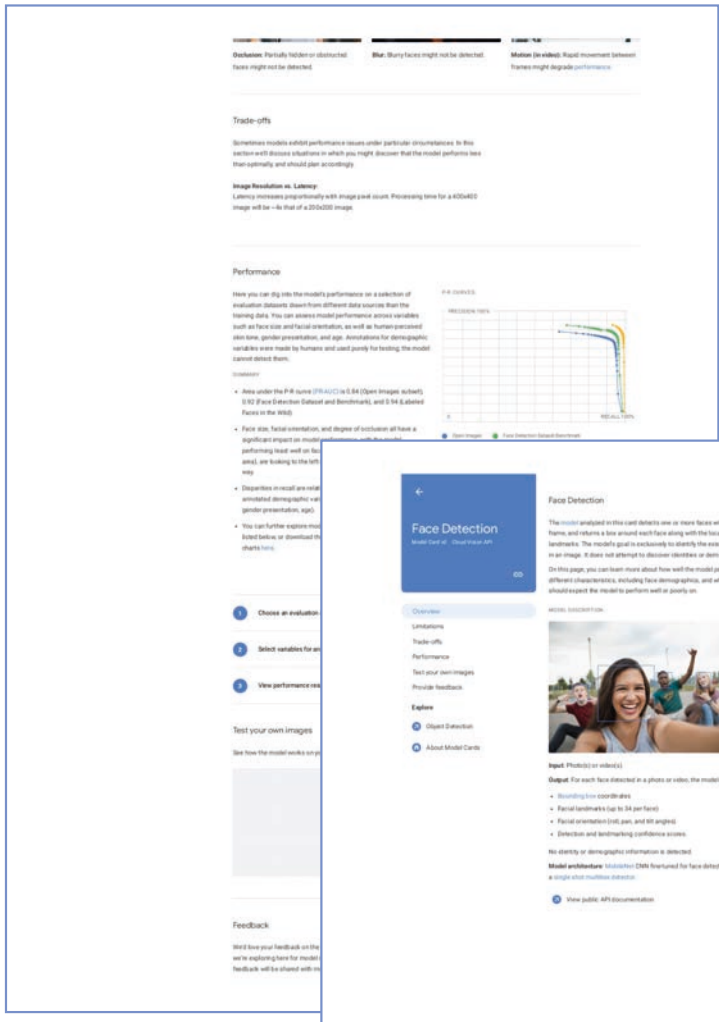

Ranking Facts
Recipe
Attribute Weight
PubCount 1.0
Faculty 1.0
GRE 1.0
Ingredients
Attribute Importance
PubCount 1.0
CSRankingAllArea 0.24
Faculty 0.12
Importance of an attribute in a ranking is quantified by the correlation coefficient between attribute values and items scores, computed by a linear regression model. Importance is high if the absolute value of the correlation coefficient is over 0.75, medium if this value falls between 0.25 and 0.75, and low otherwise.
Ingredients
Top 10:
Attribute Maximum Median Minimum
PubCount 18.3 9.6 6.2
CSRankingAllArea 13 6.5 1
Faculty 122 52.5 45
Overall:
Attribute Maximum Median Minimum
PubCount 18.3 2.9 1.4
CSRankingAllArea 48 26.0 1
Faculty 122 32.0 14
Diversity at top-10
DeptSizeBin
Large Small
Regional Code
NE W MW SA SC
Diversity overall
DeptSizeBin
Large Small
Regional Code
NE W MW SA SC
Stability
Top-K Stability
Top-10 Stable
Overall Stable
Fairness
DeptSizeBin FA*IR Pairwise Proportion
Large Fair Fair Fair
Small Unfair Unfair Unfair
A ranking is considered unfair when the p-value of the corresponding statistical test falls below 0.05.
Fairness
FA*IR Pairwise Proportion
DeptSizeBin p-value adjusted α p-value α p-value α
Large 1.0 0.87 0.98 0.05 1.0 0.05
Small 0.0 0.71 0.0 0.05 0.0 0.05
FA*IR and difference in proportions (Proportion) are measured with respect to 26 highest-scoring items (the top-K). The top-K contains 100 items or one half of the input, whichever is smaller.

Caso studio 5

TITOLO: Model Cards for Model Reporting
AUTORE: (Google) Mitchell et al.
ANNO: 2019
FORMATO: schede modulari di pagine online interattive
TARGET: stakeholders, con il fine di comparare i modelli per l'implementazione
KEYWORD: Dataset, Trasparenza, Inclusività, Valutazione dei casi d'uso

Documentazione sintetica che accompagna i dataset per chiarire i casi d'uso previsti dai modelli di IA e ridurre il loro utilizzo per contesti nei quali non sono adatti. Le schede forniscono una valutazione dei modelli di ML, in riferimento a varie condizioni: riguardo diversi gruppi culturali, demografici o fenotipici e gruppi intersezionali che sono pertinenti ai domini di applicazione previsti. Le schede rivelano anche il contesto in cui i modelli sono destinati a essere utilizzati, i dettagli delle procedure di valutazione delle prestazioni ed altre informazioni rilevanti. La struttura può essere usata per documentare qualsiasi modello di apprendimento automatico.

FIG. 36 Google Model Cards for Model Reporting screen.

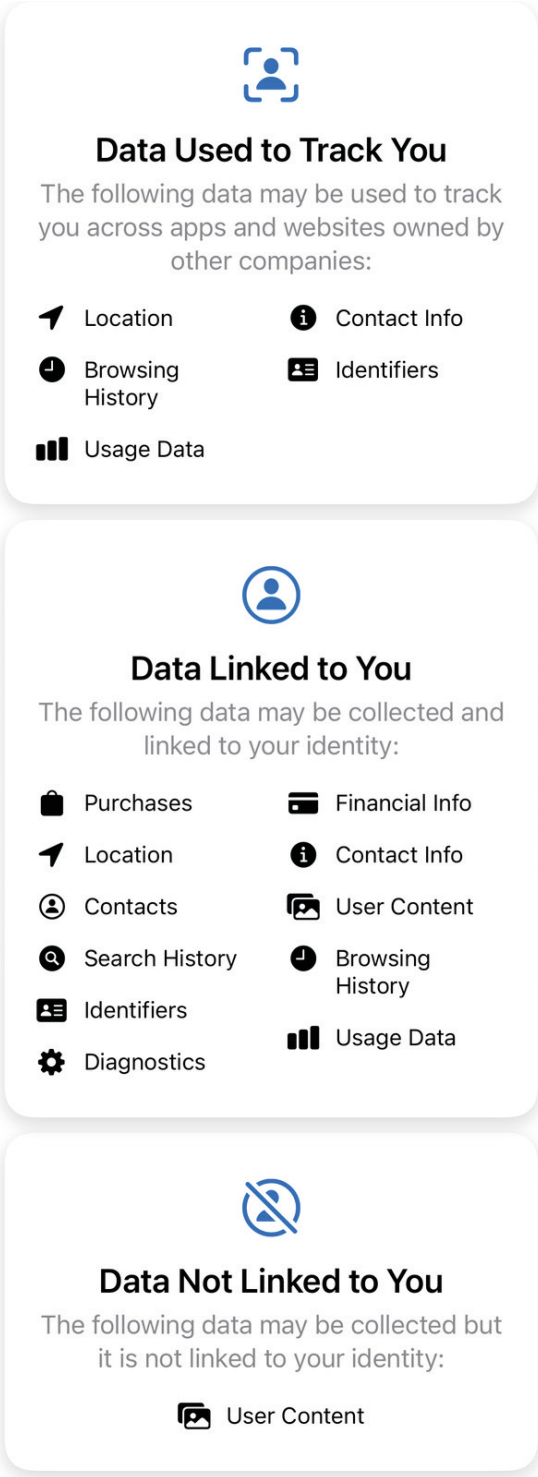

Trade-offs
Performance
Test your own images
Feedback
Face Detection
Overview
Limitations
Trade-offs
Performance
Test your own images
Provide feedback
Explore
Object Detection
About Model Cards
Face Detection
Limitations

Caso studio 6

TITOLO: Apple Privacy Nutrition Labels
AUTORE: Apple
ANNO: 2020
FORMATO: schede modulari online
TARGET: utenti delle applicazioni; sviluppatori
KEYWORD: Privacy, Dati personali

Il sito corporate Apple dichiara: «transparency is the best policy». *Apple Privacy Nutrition Labels* sono documenti che gli sviluppatori sono tenuti a compilare al rilascio di una nuova applicazione. Nei documenti sono identificati con precisione i dati ai quali l'app può avere accesso e come questi ultimi potrebbero venire utilizzati. Al momento esistono circa 34 diverse di etichette standardizzate.

FIG. 37 Apple Privacy Nutrition Labels.


Data Used to Track You
The following data may be used to track you across apps and websites owned by other companies:
Location
Contact Info
Browsing History
Identifiers
Usage Data
Data Linked to You
The following data may be collected and linked to your identity:
Purchases
Financial Info
Location
Contact Info
Contacts
User Content
Search History
Browsing History
Identifiers
Usage Data
Diagnostics
Data Not Linked to You
The following data may be collected but it is not linked to your identity:
User Content

4.3 Il focus: stereotipi su ruoli e oggetti di genere nelle immagini

> Gli uomini agiscono, le donne appaiono. Gli uomini guardano le donne. Le donne guardano se stesse mentre sono guardate. Questo determina non solamente la maggior parte delle relazioni fra uomini e donne ma anche il rapporto delle donne con se stesse. L'osservatore della donna è maschile; l'osservata femminile. Così lei si trasforma in oggetto. Più specificamente in oggetto di visione: in veduta. (Berger 1972: 49)

Come dichiara Pierre Bourdieu gli stereotipi di genere sono tra gli stereotipi sociali più potenti e anche tra i più difficili da de-costruire perché «ciò implica una denaturalizzazione delle rappresentazioni sociali e una decostruzione di questo mondo incorporato sotto forma di habitus» (Bourdieu 1998: 16).

Nessuno degli esempi in letteratura dei fogli illustrativi nell'IA si è soffermato a studiare i gender bias. Come già abbiamo visto, spesso le immagini dei dataset sono etichettate da lavoratori a cottimo (si veda Amazon Turkers) che, volenti o nolenti, codificano nelle etichette visioni stereotipiche, causa la velocità di esecuzione con cui devono portare a termine il lavoro. L'*Implicit Association Test* (IAT) dimostra proprio questo: l'utente medio agisce in modo completamente diverso se gli viene chiesto di comparare a colpo d'occhio due concetti che gli sembrano familiari, in contrasto a due concetti che appaiano differenti.

Infatti l'IAT registra che a nomi maschili viene associata la carriera, a nomi femminili la famiglia; al maschio la matematica e la scienza e alla femmina l'arte.

Molti momenti della vita sono divisi in compiti con un rapporto di dicotomia: anche i ruoli di lavoro, ad esempio, sono ancora altamente codificati.

> Esiste una segregazione sociale e culturale che discrimina gli uomini nell'ingresso in alcuni lavori considerati "da donne", come la maestra d'asilo (nido e infanzia), l'assistente sociale, l'ostetrica e l'estetista, vale a dire tutte le occupazioni non coerenti con una rappresentazione tradizionale della maschilità. Chi viola questa regola provoca reazioni negative. Il femminile, nel nostro immaginario, è ancora profondamente legato all'idea di cura - dell'infanzia, in primo luogo, ma anche del corpo e dell'aspetto - cura che può includere gli uomini, senza compromettere l'idea di maschilità, solo se altamente specializzata (pediatra, psicologo, psichiatra). (Bucchetti 2020)

I tempi evolvono ma alcuni stereotipi si fissano sempre di più. Si veda, ad esempio, il lavoro di cura in tempo di pandemia, dove «più donne si prendono cura di figli, nipoti, anziani e/o persone con disabilità ogni giorno per 1 ora o più rispetto agli uomini. La pandemia di COVID-19 ha aumentato la pressione sulle famiglie, specialmente sulle donne e sulle madri sole» (EIGE 2020; traduzione mia).

Esattamente come si insegna l'alfabeto al bambino, addestrare un modello con un particolare dataset di allenamento significa trasmettere conoscenza all'algoritmo che "nascerà". Per questo, se si parla di immagini, è necessario utilizzare segni che siano facili da imparare per il modello, facendo uso di risorse visuali disponibili per raccontare storie, definizioni e situazioni differenti. Il dataset deve trasformare azioni e oggetti opachi in forme facilmente comprensibili. Si può in questo senso riprendere in gran parte il filo logico del discorso di Goffman in *Gender Advertisement*. Sebbene l'autore parli delle immagini pubblicitarie, il concetto non si allontana poi così tanto dalle figure che ritroviamo nei dataset: immagini per lo più stock, o derivate da pubblicità poi trovate su google, o ancora immagini private ma scelte da chi ha composto il dataset per ricalcare situazioni convenzionate, o recuperate da social media quali *Flickr*.

Il design grafico e in particolar modo quello pubblicitario non crea dal nulla le espressioni ritualizzate da loro impiegate, ma si basa sullo stesso corpus di dimostrazioni e di riti che tutti noi viviamo quotidianamente, dato che corrisponde alla risorsa di tutti noi che partecipiamo alle si-

tuazioni sociali e con lo stesso fine: rendere le azioni su cui posiamo lo sguardo intellegibili. Se non altro, i pubblicitari convenzionalizzano le nostre convenzioni, stilizzano ciò che è già una stilizzazione, fanno un uso di *ciò che è già qualcosa*. Le campagne pubblicitarie consistono in iper-ritualizzazioni (cfr. Goffman 2015: 133).

La ricerca affonda le radici in due studi scientifici: *Men Also Like Shopping* e *Women also Snowboard*. Questi non solo testimoniano la presenza di bias all'interno dei dataset di allenamento, seguendo uno studio puramente matematico, ma affermano anche che l'allenamento del modello su questi dati amplifica ulteriormente il bias. Da questi studi si possono inoltre cogliere vari pattern che si verificano quasi indistintamente in tutti i dataset di immagini.

> One common theme is the notion of bias amplification, in which bias is not only learned, but amplified. For example, in the image captioning scenario, if 70% of images with umbrellas include a woman and 30% include a man, at test time the model might amplify this bias to 85% and 15%. (Burns et al. 2019)

La prima ricerca *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2017) studia i dati e i modelli di predizione strutturati in due ambiti: nella classificazione "multilabel" di oggetti e nell'etichettatura visiva dei ruoli semantici. La ricerca giunge alla conclusione che i dataset contengono significative distorsioni di genere e i modelli addestrati su questi dataset amplificano le distorsioni esistenti. Ad esempio, l'attività "cooking" ha più del 33% di probabilità di coinvolgere attori di sesso femminile rispetto a quelli di sesso maschile ed un modello addestrato su quel dataset amplifica ulteriormente la disparità al 68%. La soluzione proposta dei ricercatori consiste nel porre vincoli a livello di corpus, per riuscire a calibrare i modelli di previsione e di progettare un algoritmo basato su «lagrangian relaxation for collective inference».

L'etichettatura visiva del ruolo semantico è stata analizzata nel dataset imSitu. I verbi del dataset sono fortemente sbilanciati verso agenti di sesso maschile: il 64,6% dei

verbi privilegiano infatti uomini, con una distorsione media di 0,7 - ossia di 3 a 1.

Ananlisi del bias in ImSitu vSRL

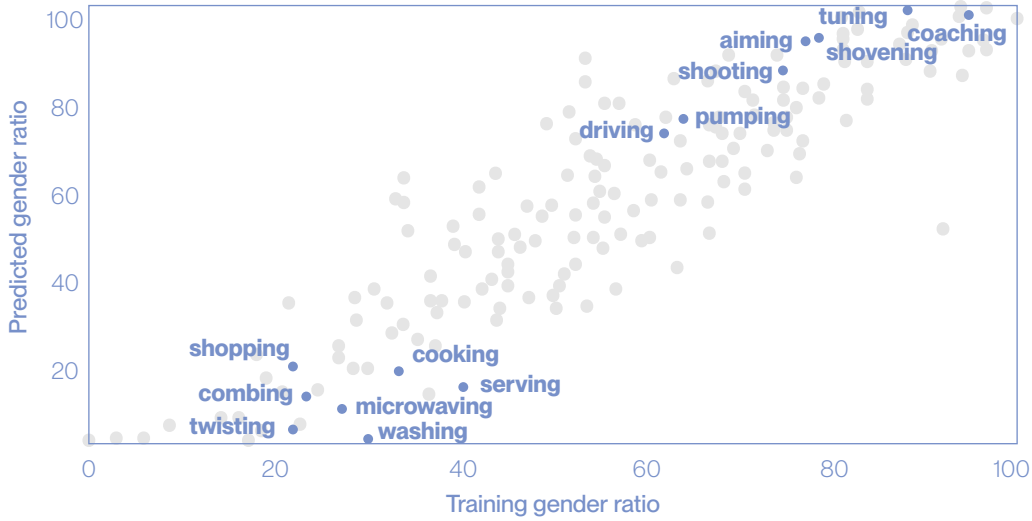


FIG. 38 Analisi del bias in imSitu. Immagine adattata da *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2017).

Nella figura 38 vengono riportate nell'asse x diverse etichette di attività che rivelano pregiudizi problematici. Ad esempio *shopping*, *cooking*, *serving*, *combing*, *microwaving* sono orientati verso un agente di sesso femminile, mentre *driving*, *pumping*, *shooting*, *coaching* verso un agente di sesso maschile. Lo studio svolge una previsione su come il modello si comporterà con un dataset inedito, non conosciuto. Scopre che l'addestramento su imSitu amplifica il bias (visibile nella figura 39 lungo l'asse y). L'amplificazione del bias è dello 0,05 in media, dove il 45,75% dei verbi mostra tale amplificazione. I verbi che erano inizialmente distorti tendono ad avere un'amplificazione più forte: i verbi con bias di allenamento superiore a 0,7 hanno un'amplificazione media di 0,07. Ad esempio *serving*, che aveva solo una piccola polarizzazione dello 0,4 verso il sesso femminile nel set di allenamento, è ora fortemente polarizzato verso il sesso femminile con un tasso di 0,12. Il verbo *tuning*, originariamente orientato verso il sesso maschile per il 0,87, ora ha esclusivamente agenti maschili.

La classificazione multilabel degli oggetti è valutata nel dataset MS-COCO. Nella figura 39 lungo l'asse delle x, si analizza l'orientamento degli oggetti in MS-COCO rispetto al genere. L'86,6% degli oggetti nel dataset è sbilanciato verso il sesso maschile, con una media dello 0,65.

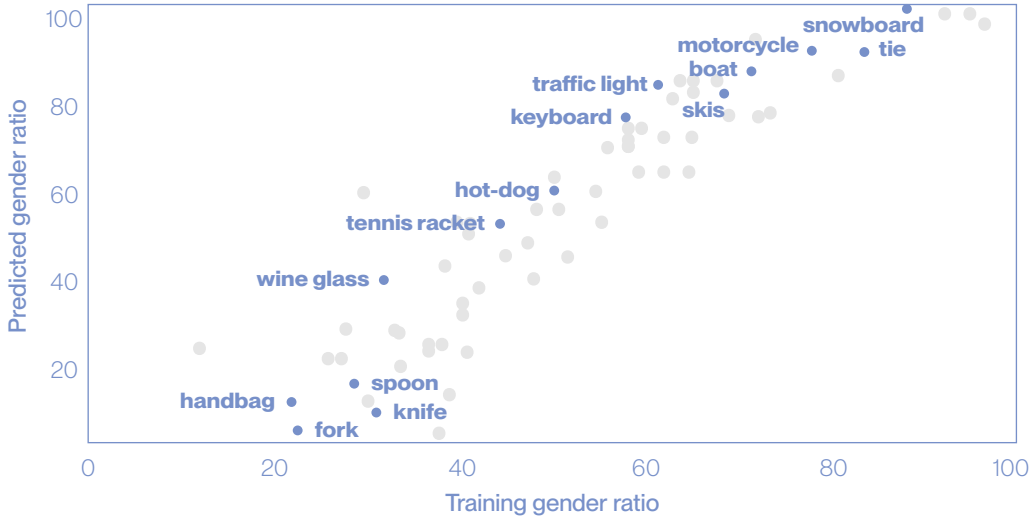
La figura 39, lungo l'asse y, mostra il tasso di uomini (% di entrambi i generi) nelle previsioni su un dataset inedito. L'amplificazione media del bias su tutti gli oggetti è dello 0,03, dove il 65,67% dei sostantivi mostrano una amplificazione. Gli oggetti con bias di allenamento superiori a 0,7 hanno un'amplificazione media di 0,08. Ad esempio, gli oggetti della cucina, già precedentemente orientati verso il sesso femminile, quali *knife fork* o *spoon*, sono tutti amplificati, mentre le categorie tecnologiche quale *keyboard* e *mouse* hanno aumentato il loro bias maschile di oltre lo 0,1.

Con questi due studi i ricercatori confermano le due ipotesi iniziali che:

1 sia il dataset ImSitu che MS-COCO sono influenzati dal genere nel numero di immagini per verbo o oggetto;

FIG. 39 Analisi del bias in MS COCO. Immagine adattata da *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2017).

Analisi del bias in MS COCO MLC



2 i modelli addestrati su questi dataset amplificano il pregiudizio di genere esistente;

3 il grado di amplificazione del bias è correlato al grado del pregiudizio iniziale.

Il secondo report *Women also Snowboard: Overcoming Bias in Captioning Models* (Burns et al. 2019), analizza la generazione di tag descrittivi gender-specific (ad esempio uomo, donna) basate sull'aspetto della persona o sul contesto dell'immagine. La ricerca introduce un modello, denominato Equalizer, che promuove un'uguale probabilità di genere quando non ci sono evidenze di genere in un'immagine. Il modello deve specificatamente osservare la persona, piuttosato che cercare spunti di contesto per generare la predizione del genere. La ricerca prende in analisi nuovamente il dataset MS COCO (facendo esplicito riferimento alla ricerca sopracitata) con il fine di valutare l'amplificazione dei bias nella predizione strutturata. Individua un rapporto di foto con soggetti di tag genere femminile/maschile di circa 1:3 nel dataset.

Nella tabella sottostante i ricercatori hanno riportato la percentuale di immagini in cui il genere è previsto correttamente o erroneamente e quando non viene generata alcuna parola specifica di genere (*other*). In tutti i modelli da loro studiati, l'errore per "Men" è abbastanza basso (4,23%), mentre è di molto maggiore l'errore per la classe di minoranza "Women" (34,11%).

FIG. 40 Analisi della correttezza dell'etichettatura di immagini maschili e femminili. Immagine adattata da *Women also Snowboard: Overcoming Bias in Captioning Models* (Burns et al. 2019).

	Women	Men
Correct	46,28	75,05
Incorrect	34,11	4,23
Other	19,61	20,72

Per quanto riguarda gli oggetti ci sono ampie differenze tra i generi, ad esempio “umbrella”, “kitchen”, “cell phone”, “table” e “food” sono altamente legati a tag che identificano individui di sesso femminile, mentre “skateboard”, “baseball”, “tie”, “motorcycle” e “snowboard” sono legati a tag che identificano individui di sesso maschile.

Un caso studio particolarmente interessante è la parola "kitchen" in cui il rapporto di genere donna-uomo è 0,946. Considerato che il dataset contiene un rapporto di genere donna-uomo di circa 1:3, un rapporto di genere vicino a 1,0 per un oggetto specifico suggerisce che una proporzione maggiore di immagini "woman" include una "kitchen" rispetto alle immagini "man". I ricercatori ipotizzano che molti errori di classificazione si verifichino a causa del fatto che il modello prevede il genere sulla base di prove visive sbagliate. Per questo affermano che il modello debba essere “right for the right reason”, quindi deve esprimere un giudizio corretto perché si basa sulle giuste prove.

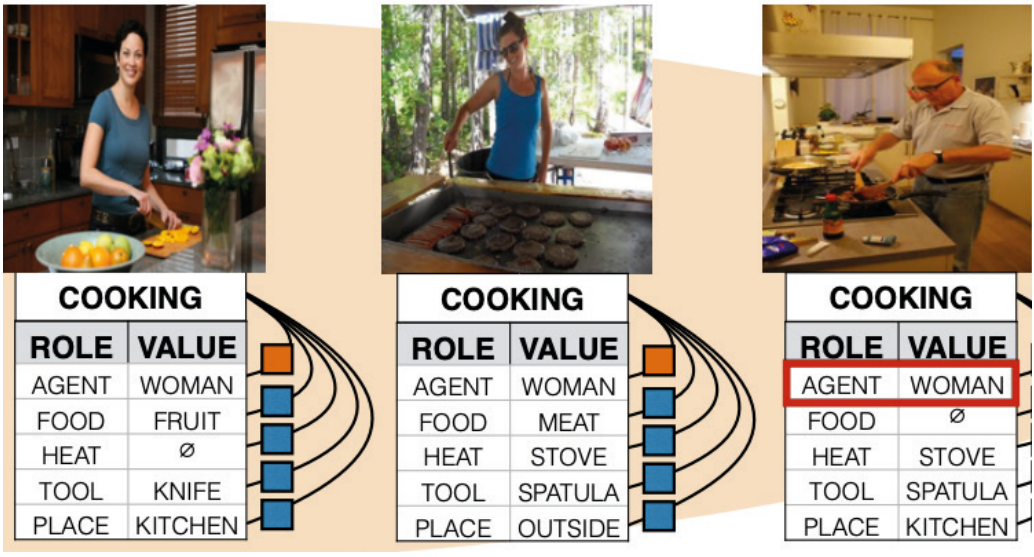


FIG. 41 Screen dal dataset imSitu. Nel set di allenamento solo il 33% delle immagini in cucina ha come agente un uomo. Dopo aver addestrato un *Conditional Random Field*, il bias è amplificato: l’uomo è agente per il 16% delle immagini di cucina. Immagine adattata da *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2019).

4.3.1 Metodi di ricerca visuale: casi studio

Non esiste, fino a prova contraria, una letteratura che spieghi in che modo analizzare le immagini dei dataset dal punto di vista qualitativo. È risultato quindi necessario ricorrere a studi paralleli, che analizzano le immagini pubblicitarie, dei social media, stock o amatoriali.

Il primo passo è stato quindi quello di effettuare una raccolta degli stereotipi, dei segni e dei significati reiterati per punti, a valle della ricerca esplorativa illustrata nel paragrafo 3.3 con l'osservazione delle figure stereotipiche nei sistemi algoritmici di riconoscimento visivo. Il passaggio che segue è stata la *schedatura*: «mediante la decostruzione e successiva ricomposizione degli stereotipi, sperimentando linguaggi principalmente grafici e fotografici, messi a sistema tramite una scheda» (Tolino 2012).

Sono stati quindi ricercati i lavori di raccolta di immagini per studi e azioni di denuncia di stereotipi di genere. In particolare, gli stereotipi che si analizzeranno nei dataset si basano sulla deformazione dei ruoli che, come afferma Lukas in *Gender Ads*[3]:

> is exhibited in many situations in popular culture and advertising. The roles offered to women are very limited. Many involve a woman's confinement to the domestic sphere—caring for children, cleaning the house, shopping for groceries and making meals for a husband. Nancy Chodorow makes the point that the maintenance of gender subordination in world cultures is very much due to the universal functions and values of the family. (Lukas 2002)

3 Gender Ads è un progetto di un sito web che cataloga tutti i tipi di pubblicità che coinvolgono il genere.

Il paragrafo cerca di riassumere tutti i tipi di pattern che potrebbero occorrere nelle immagini dei dataset, sia per le figure femminili che per quelle maschili, con il fine ultimo di riconoscere caratteri e codici ricorrenti nelle numerose rappresentazioni di genere presenti nei diversi media.

Goffman, analizzando nel 1979 il rapporto uomo-donna nella pubblicità cartacea, sostiene che si assiste ad uno sbilanciamento nei confronti dell'uomo: la donna

appare sempre in una *posizione subordinata* (cfr. Goffman 1979). La subordinazione è un tema ripreso anche da Caratti, la quale afferma che risulta espressa nei ruoli che la figura riveste «conformi al binomio seduzione-riproduzione» (Caratti 2015). Non solo, risulta espressa anche, in modo meno esplicito, attraverso «le forme della rappresentazione, dove la subordinazione è evidente a partire dall'analisi delle forme espressive e dai processi di significazione: ad esempio la scelta e riproposizione di alcuni canoni estetici di rappresentazione nella composizione dell'immagine (le inquadrature, i colori, le luci)» (*ibidem*).

«Goffman definisce il fenomeno come comportamento non verbale ritualizzato indicato dalla prossemica attraverso l'osservazione della postura, delle espressioni, dei gesti e delle relazioni spaziali tra i soggetti maschili e femminili in presenza o supposti» (*ivi*: 100). Goffman individua come categoria ricorrente la *Relative Size* (o grandezza relativa): ossia la connessione tra il peso sociale (il potere, l'autorità, il rango, l'ufficio, la fama) e lo spazio che le figure occupano nella composizione. Gli uomini, secondo la ricerca di Courtney e Lockeretz (1979), oltre ad essere più alti, vengono non di rado mostrati mentre considerano le donne alla stregua di oggetti sessuali o come mere coadiuvanti domestici e, per rafforzare il ruolo dominante, spesso sono in piedi o seduti più in alto delle donne (Goffman 1979). Goffman richiama questo stereotipo nella categoria *Ritualization of Subordination* (o ritualizzazione della subordinazione): un modello ricorrente di deferenza è quello di chinarsi fisicamente. Lo stesso effetto si verifica quando consideriamo lo spazio topologico e le posizioni delle figure nella composizione, ossia quando la figura femminile viene collocata "dietro" a quella maschile.

> According to Fowles (1996), author of Advertising and Popular Culture, women are often portrayed as passive creatures in advertisements, being placed indoors and often in a reclining position or sitting. Masse & Rosenblum's (1988) study of the portrayal of men and women in magazine images supports this, finding that females are more likely to be depicted in partial views and/or in a subordinate position (27% for females versus 4% for males). (Mason 2003)

FIG. 42 Lora Lamm manifesto *Vivere sui Tappeti* per la Rinascente di Milano (1959).

Un esempio che si oppone con forza a questa visione sono le composizioni di Lora Lamm, della serie *Vivere sui tappeti*. In particolare, se si esamina *Cento tappeti nuovi per Milano* (1959), si nota come la composizione si sviluppi a partire da una visione in cui entrano in gioco sia la visione zenitale sia la visione frontale (alla quale è soggetta il gatto) che, incrociati, producono un effetto di immediata reciprocità. L'uomo entra nel campo verticalmente dall'alto a sinistra, la donna dal basso a destra, creando un equilibrio perfetto e simboleggiando la medesima dignità ed il medesimo ruolo delle figure in scena (cfr. Bucchetti 2021). La traduzione visuale operata dal designer può contribuire alla costruzione della rappresentazione di una dimensione sociale equa.

Molti spot e pubblicità strutturano la femminilità come una forma di *pura sessualità* - come se ogni donna fosse ninfomane (cfr. Lukas 2002). Sullivan e O'Connor (1988) affermano che le donne nelle pubblicità hanno principalmente due ruoli: quello decorativo e quello del personaggio seducente e sexy (cfr. Caratti 2015: 33). La donna moderna «viene dipinta come una "donna soggetto", auto

affermativa e sicura di sé, ma il principale potere che le viene concesso è quello di sedurre. Una seduzione che avviene soprattutto tramite il corpo, da mantenere paradossalmente "per sempre" *giovane e attraente*» (*ibidem*; corsivo mio). Si veda l'eclatante esempio della pubblicità *Carl's Natural Burger*, dove il tema della bellezza e della sensualità della donna è centrale. Kang (1997) suggerisce la categoria di analisi *Body Display* (esposizione del corpo), secondo la quale le modelle delle pubblicità nelle riviste mostrano un alto grado di nudità, che consiste in un fattore rilevante di stereotipizzazione che sottolinea nuovamente l'aspetto della seduzione. I vestiti "rivelanti" per l'autore includono: minigonne, gonne strette o abiti da sera che espongono la scollatura, pantaloncini corti, vestiti *see-through*, costumi da bagno, lingerie, asciugamani. Gli scatti ravvicinati in cui le spalle delle modelle sono nude sono considerati nudità.

La stessa Caratti conferma il ritratto distorto del femminile, portatore di una *cultura «ipersessuale»* che riconferma a livello simbolico i rapporti di dominanza uomo/donna:

> Ricerche sulla disuguaglianza di genere anni Ottanta evidenziano come i media presentino con maggior frequenza visi maschili e corpi femminili, associando quindi ai maschi qualità intellettuali e alle femmine qualità fisiche ed emotive (cfr. Archer, Iritani, Kimes, Barrios 1983). Secondo queste indagini le persone ritratte con un focus sul volto sono giudicate più intelligenti (fenomeno del "face-ism"); le donne al contrario tendono ad essere "smembrate", rappresentate con parti del corpo, e in particolare con gli attributi sessuali primari e secondari (fenomeno del "body-ism"). (Caratti 2015: 100)

«[...] se vediamo una figura intera tendiamo a concentrarci sugli attributi fisici, come l'avvenenza o la sensualità; se prevale il volto emergono invece i meriti intellettuali, la personalità e i valori emotivi» (Falcinelli 2020:443).

Falcinelli parla però anche del nudo che, soprattutto in pittura, coinvolge anche corpi maschili, ma in pose molto differenti dai corpi femminili:

> [...] se diamo una scorsa alla pittura di nudo, ci accorgiamo

> che la maggioranza dei corpi maschili è raffigurata in piedi, mentre quelli femminili sono sdraiati, e quindi spesso i primi occupano quadri verticali, le seconde, orizzontali. [...] mentre al nudo virile si chiede di rappresentare valori di forza ed eroismo con cui identificarsi, quello femminile è sentito come preda, come oggetto sessuale già pronto per essere consumato. Stare in piedi pone il protagonista in un ruolo attivo, stare stesi, passivo, in senso letterale e metaforico. (Falcinelli 2020:95)

Function Ranking (o classifica delle funzioni) è una terza categoria di Goffman che evidenzia come, quando un uomo e una donna collaborano, è probabile che l'uomo svolga il ruolo esecutivo. Questa gerarchia di funzioni è raffigurata sia all'interno di un ambiente lavorativo sia nella quotidianità. Courtney e Lockeretz (1979) sostengono il medesimo concetto, specificando che solitamente le donne non sono mostrate come professioniste, al contrario della controparte maschile: «gli uomini agiscono e le donne appaiono» (Berger 2008). La nozione è comprovato dagli studi analizzati nel paragrafo 3.3, dove quasi la maggior parte delle immagini dei soggetti femminili risultano inappropriate: indipendentemente dal settore occupazionale, sono infatti raffigurate in modo sessualizzato e non professionale. Non esistono paragoni per i soggetti maschili.

Un tropo ricorrente consiste nell'*ambientazione*: le donne sono raramente mostrare fuori casa e sono solitamente ricondotte ad *ambienti interni e casalinghi* (cfr. Courtney e Lockeretz 1979; Caratti 2015). Come parte della loro connessione alla sfera privata, le donne sono spesso associate allo *shopping* e presentate come inseparabili dai loro beni di consumo (cfr. Lukas 2002).

Un'altra sottocategoria dell'ambientazione casalinga è la donna vista come madre, c'è infatti una suggestione ideologica che le donne siano associate alla loro procreatività e al loro status percepito di madri (cfr. Lukas 2002). Ma, come sottolinea Bucchetti:

> si è di fronte a una dicotomia tra la rappresentazione della donna come casalinga, nell'atto di compiere i lavori domestici, e quella di oggetto sessuale del desidero maschile, rive-

> stimento erotico dei beni di consumo, per arrivare alla quadratura del cerchio, attraverso la rappresentazione di una donna impegnata nei propri lavori domestici ma non privata della propria seduttività e attrattività. (Bucchetti 2012: 91)

In questo tipo di raffigurazioni è possibile riprendere la quarta categoria di analisi di Goffman, ossia il *Feminine Touch* (o tocco femminile): le donne, più degli uomini, sono raffigurate mentre usano le dita e le mani per tracciare i contorni di un oggetto, cullarlo o accarezzarne la superficie. Questo tocco rituale è ben distinto dal tocco funzionale che afferra, manovra o stringe. Il concetto è ripreso da Bucchetti (2012), la quale afferma che spesso le mani femminili sono il focus dell'immagine pubblicitaria, sono una «porzione privilegiata del corpo per testimoniare il lavoro, l'efficacia del risultato, la semplicità e la consuetudine del gesto».

Le donne sono poi solitamente mostrate come molto *emotive*, caratterizzate da stati alterati di coscienza innescati da prodotti di uso comune. Il loro linguaggio corporeo esprime un'età mentale imprigionata nella prima adolescenza, sottolineato dalla posa delle mani, dall'inclinazione del collo, dalla postura delle gambe. Nelle pubblicità di sconti, offerte o promozioni le donne sono stupite, con la bocca aperta e gli occhi sbarrati ad evidenziarne la sorpresa. Qui si sottolinea il tropo della donna *trasecolata*, esempio di rappresentazione in grado di svilire "silenziosamente" le donne: rappresenta un caso di particolare interesse sia per la mancanza di referenzialità sia per l'inspiegabilità della sua diffusione. I criteri che definiscono questa posa sono: l'aspetto giovane e gradevole, l'inquadratura ripresa da un punto di vista preferibilmente frontale-parallelo, gli occhi spalancati, lo sguardo sgranato, la bocca aperta e, accanto alla bocca, la posizione delle mani costituisce un dettaglio fondamentale. Questa rappresentazione declina lo stupore mettendo l'accento su un sentimento che sconfina nell'incredulità e nello smarrimento, nel pudore verso ciò che si prova, quasi a volersi schermire, nell'esaltazione che restituisce in figura l'esclamazione "WOW" collegata a un fatto inaspettato, nell'incontenibilità dell'emozione che scoppia dentro al soggetto (cfr. Bucchetti 2021).

Goffman studia una categoria che similmente raffigura le donne impegnate in coinvolgimenti che le rimuovono psicologicamente dalla situazione sociale, lasciandole disorientate e dipendenti dalla protezione di altri eventuali soggetti presenti. Questa categoria è chiamata *licensed withdrawal* (ritiro autorizzato). Si verifica ad esempio, nel caso di una figura che guarda altrove, o che intrattiene una conversazione telefonica (cfr. Goffman 1979).

L'*esotizzazione* delle donne avviene in molti livelli nella cultura popolare e consiste nel presentarle in contesti particolarmente strani o surreali: donne con animali, in situazioni bizzarre, creature mitiche o altri scenari irreali.

Questo tropo fa emergere il concetto di intersezionalità - l'idea di Kimberle Williams Crenshaw di come le donne di colore e le donne di classe inferiore siano colpite da maggiori forme di oppressione istituzionale e individualizzata. Crenshaw offre inoltre un'analisi di una serie di vecchi film, come ad esempio *Wild at Heart*, dove dimostra che le donne di colore sono ritratte spesso come sacerdotesse voodoo, mostri e simili. Hill Collins (1990) riprende l'argomento affermando che le donne di colore sono ritratte nei media come personaggi docili, dominatrici, irresponsabili e promiscue se afroamericane, pigre o civettuole se latine, seducenti principesse se native americane (cfr. Lukas 2002).

1

2


Picture you with a kitchen extension phone, enjoying its convenience and helpfulness. And picture the chops that won't burn, the puddings that won't boil over, because you'll never need to leave the kitchen to make or take a call. In fact, why stop with picturing when ordering's so easy and inexpensive? Just call the Business Office or ask your telephone man, and take your choice of kitchen phones in attractive styles and colors.
BELL TELEPHONE SYSTEM
(1963)

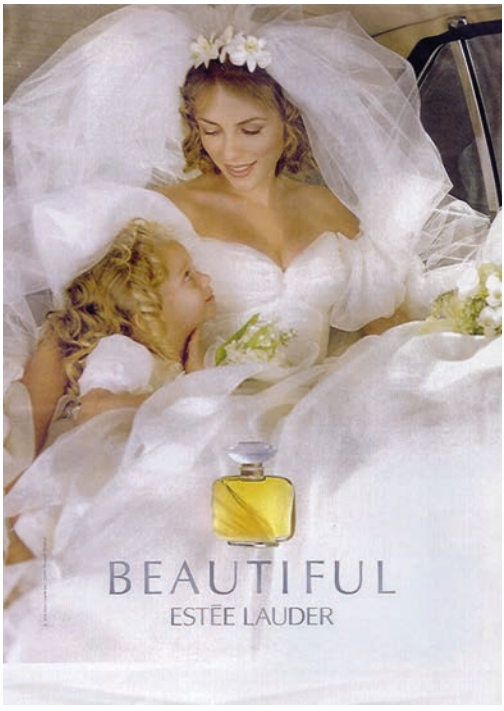
3


guess eyewear
guess eyewear

4

5

6



7

8

FIG. 43
1. Rosa e blu: Jeong Mee Yoon The Pink & Blue Project: (2005-attuale).
2. Oggettificazione sessuale: Carl's jr pubblicità per il Super Bowl (2015).
3. Licensed Withdrawal: Bell Telephone System (1963).
4. Donna sessualizzata mentre svolge faccende domestiche: Guess Eyewear (1990).
5. Body display, dimensione relativa e function raking: pubblicità del marchio Tom Ford (2008).
6. Relative Size. pubblicità del marchio Gucci, in Cosmopolitan Magazine.
7. Donna come madre: Estee Lauder Magazine annuncio pubblicitario (1999, UK)
8. Ritualization of subordination: pubblicità Gucci, in Femina Magazine (2003).

5.0 Fogli illustrativi per l'analisi di dataset di immagini

With big data comes big responsibilities.
Kate Crawford 2013

L'articolo 14 della Convenzione *Europea sui Diritti Umani* proibisce la discriminazione, in quanto afferma che:

> Il godimento dei diritti e delle libertà enunciati nella presente Convenzione deve essere assicurato senza alcuna discriminazione per ragioni di sesso, razza, colore, lingua, religione, opinione politica o di altro genere, origine nazionale o sociale, associazione a una minoranza nazionale, proprietà, nascita o altra condizione.[1]

La Convenzione specifica che la discriminazione proibita può essere sia *diretta* che *indiretta*. La prima indica un tipo di discriminazione basata su una caratteristica della persona, quale l'etnia o il sesso, mentre la seconda una pratica che a prima vista appare neutrale ma in realtà coinvolge aspetti di discriminazioni contro un gruppo. Il processo decisionale dell'IA può portare alla discriminazione indiretta involontaria. La legge sulla non discriminazione può essere utilizzata per combattere le decisioni discriminatorie dell'IA, ma ha molteplici punti deboli, tra i quali la necessità della dimostrazione che una pratica colpisce in modo sproporzinoato un gruppo protetto.

La legge sulla protezione dei dati (GDPR) potrebbe aiutare a mitigare i rischi di discriminazione, in quanto richiede trasparenza e obbliga alla compilazione di una nota della privacy per tutti le fasi di un processo decisionale che coinvolga i dati personali. In determinate circostanze, il GDPR e la *Convenzione sulla protezione dei dati* richiedono alle organizzazioni di condurre una valutazione d'impatto sulla protezione dei dati (DPIA). La DPIA è richiesta quando una pratica potrebbe comportare un rischio elevato per i diritti e le libertà delle persone fisiche, ad esempio vengono prese decisioni completamente automatizzate con effetti legali a discapito delle persone. Il rischio di discriminazione ingiusta o illegale deve però essere considerato anche quando si conduce una DPIA: molti tipi di decisioni automatizzate rimangono fuori dal campo di applicazione delle norme del GDPR. La disposizione del GDPR si applica infatti solo alle decisioni basate *esclusivamente* sull'automatizzazione e per questo motivo, quando un impiegato di una banca nega un prestito sulla base di una raccomandazione di un sistema

1 Articolo 14: *Divieto di discriminazione.* 1994. Articolo 14, vol. European Court of Human Rights Council of Europe. <https://presidenza.governo.it/CONTENZIOSO/contenzioso_europeo/documentazione/Convention_ITA.pdf>.

di IA, la disposizione non si applica. La legge attuale ha dei punti deboli quando viene applicata all'IA, per questo è necessaria un'ulteriore regolamentazione per proteggere le persone dalla discriminazione illegale e dalla differenziazione ingiusta (Borgesius 2018).

Il secondo punto del report *AI and Gender: Four Proposals for Future Research* di Collet e Dillon si sofferma sull'importanza di trarre linee guida per i dati e per gli algoritmi, afferma infatti la fondamentale l'importanza di formulare linee guida specifiche per la ricerca e lo sviluppo dell'IA. Queste linee guida comprendono le definizioni di "fairness" e "bias" e, con il fine di renderle quanto più specifiche possibili, le linee guida «would be context-specific, addressing crime and policing, health, and financial services, as well as other sectors which evidently have problems with gender equality when it comes to their datasets». Collet e Dillon aggiungono che: «context-specific guidelines for data which focuses specifically on issues of gender equality, would avoid the "one-size-fits-all" approach which generalises and does not tackle issues specific to particular AI systems» (Collet e Dillon 2019: 23).

Sono apparsi negli ultimi anni un grande numero di *framework* etici, basati in modo specifico sull'IA. Il gruppo *AI4People*, guidato da Luciano Floridi, analizzando un corpus di sei framework etici sull'IA, ha riscontrato che esistono quattro principi etici imprescindibili che ogni trattato nomina. Questi principi esprimono l'essenza delle documentazioni e sono: *beneficence*, *non-maleficence*, *justice*, and *autonomy*. Nel report di Deloitte *Human values in the loop: Design principles for ethical AI* (2020), gli autori propongono impatto, giustizia e autonomia come tre principi utili a guidare le discussioni sulle implicazioni etiche dell'IA.

1 *Impatto* indica che la qualità morale di una tecnologia dipende dalle sue conseguenze: i rischi e benefici devono essere soppesati. Due principi etici ampiamente riconosciuti sono la benevolenza e la non-maleficenza. Il primo sostiene che l'IA dovrebbe essere progettata per aiutare a promuovere il benessere delle persone e del pianeta e il secon-

do prescrive che l'IA dovrebbe evitare di causare danni intenzionali o involontari;

2 *Giustizia* sostiene che le persone dovrebbero essere trattate in modo equo. È utile distinguere tra i concetti di equità procedurale e distributiva. Una politica (o un algoritmo) si dice essere proceduralmente equa se è equa indipendentemente dai risultati che produce, si dice invece essere distributivamente equa se produce risultati equi;

3 *Autonomia* esprime la necessità di prendere decisioni proprie, senza forze esterne manipolatrici. Molti dei principi appartenenti al discorso di etica nell'IA, quali la trasparenza, la spiegabilità, la privacy ecc. possono essere visti come aspetti appartenenti all'autonomia (cfr. Guszcza et al. 2020).

Sono stati considerati metodi scientifici e tecnici per arrivare alla progettazione di modelli matematici "fair" ma, come affermano Collet e Dillon, nel report *AI and Gender: Four Proposals for Future Research*, se non vengono considerati i contesti sociali e politici, le formule matematiche mancheranno dei fattori chiave per risolvere realmente il problema alla radice.

> Broadening perspectives and expanding research into AI fairness and bias beyond the merely mathematical is critical to ensuring we are capable of addressing the core issues and moving the focus from parity to justice. (Whittaker, Crawford, Dobbe et al., 2018: 8)

In questo caso quindi, «definitions of fairness would benefit from considerations of current and historical gender discrimination» (Collet e Dillon 2019: 20).

> Il tema della rappresentazione della donna, così come i media la impongono, con le implicazioni che questo modello rappresentato porta con sé è, pertanto, un tema di cui il Design della comunicazione deve farsi carico, intorno al quale è opportuno sviluppare strumenti in grado di produrre anticorpi, per acquisire - ma anche assimilare - una dimensione

> critica che renda capaci di guardare le immagini per il loro portato complessivo e per ciò che sottendono [...]. (Baule e Bucchetti 2013)

5.1 Foglio illustrativo: cos'è e come viene concepito

Per risolvere il problema del gender bias negli algoritmi di riconoscimento visivo si riparte dall'origine, dal p*eccato originale*: lo scopo è interrogare i dataset alla ricerca di stereotipi di genere. Il progetto si ispira dal gruppo DNL, il quale afferma che:

> We believe that algorithm developers want to build responsible and smart statistical models, but that there is a key step missing in the standard way these models are built. This step is to interrogate the dataset for a variety of imbalances or problems it could have and ascertain if it is the right dataset for the model. Similar to the FDA's nutrition label for food, the Dataset Nutrition Label aims to highlight the key ingredients in a dataset in addition to qualitative information that describes the dataset and its composition, collection, and management. The Dataset Nutrition Label also includes Alerts about the dataset that are relevant for particular intended

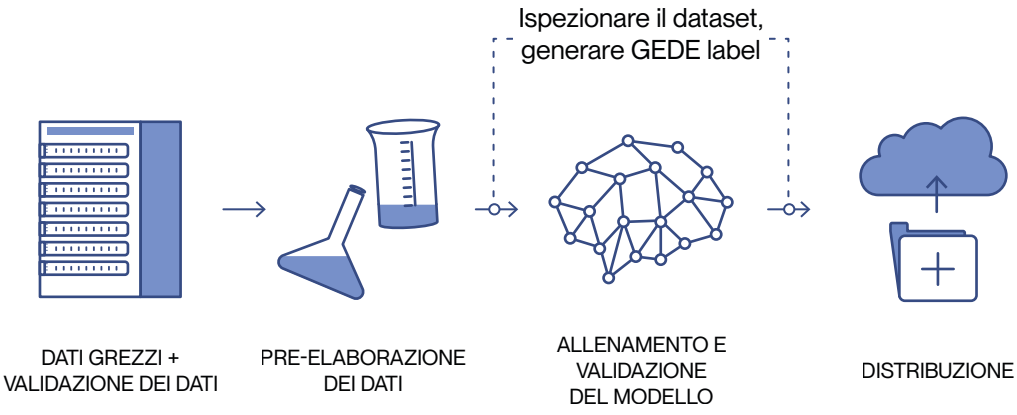


FIG. 44 Spazio di collocamento di GEDE all'interno del processo di sviluppo di un algoritmo.

> modeling objectives. Data scientists can leverage the Dataset Nutrition Label to make better, informed decisions about which datasets to use for their specific use cases, thus driving better statistical models and artificial intelligence. (Chmielinski et al. 2020; corsivo mio)

Ripartire dall'origine, quindi, significa investigare la qualità del dataset e generare delle "nutrition label". La creazione della label si inserisce nel processo di pipeline di ML tra la creazione del dataset e la pre-elaborazione del dataset, cui segue lo sviluppo del modello e la successiva distribuzione di quest'ultimo.

Il designer della comunicazione interviene in due step, per rivolgersi a due tipi diversi di destinatari: il primo consiste in modelli di ML/DL, il secondo in creatori e/o utilizzatori dei dataset. Il primo step è l'analisi visiva, che studia quali contenuti vengono trasmessi dai dataset di allenamento ai modelli di ML/DL, da cui l'algoritmo impara a riconoscere i pattern. Il secondo step è la successiva comunicazione di questa analisi in un prodotto leggibile ed utilizzabile, che offre sia una metodologia di ricerca che una visione nello specifico del gender bias nelle IA.

Occorre che l'output progettuale sia composto da un modello semplice che raccolga informazioni complesse,

FIG. 45 Doppia interazione di GEDE: modello algoritmico e utilizzatori del dataset.

che sia aggiornabile e comprensibile. I principali utilizzatori individuati sono:

- creatori del dataset, per comunicare problemi caratteristiche e la composizione del dataset;
- utilizzatori del dataset, per una scelta più consapevole;
- giornalisti e analisti dell'industria, i quali potrebbero fornire intuizioni che renderebbero più facile spiegare una tecnologia complessa a un pubblico generale;
- gruppi di difesa (*advocacy*) per comprendere meglio l'impatto dell'IA.

Il foglio illustrativo consiste in un sistema metodico di raccolta di informazioni, identificabile come etichetta nutrizionale, carta di identità, foglio illustrativo o bugiardino. Che si tratti di conoscere il contenuto del cibo, le condizioni delle strade o le avvertenze di un farmaco, ci affidiamo a informazioni ufficiali per prendere decisioni responsabili, mentre i dataset di allenamento per modelli di apprendimento automatico sono spesso distribuiti senza una chiara comprensione del loro contenuto.

Il progetto prende il nome di GEDE*: Gender Debiaser label*. Consiste in un modulo di analisi aggiuntivo per i nutrition labels "standard", che propone una metodologia di indagine e di valutazione dei gender bias nei dataset, dal punto di vista del designer della comunicazione.

Le cosiddette *statistiche di genere* sono strumenti utilizzati nell'ambito della ricerca sociale che permettono di organizzare le informazioni perseguendo vari obiettivi: l'evidenza del dato statistico relativo al sesso è risconosciuta come indispensabile per costruire argomentazioni e monitorare in modo oggettivo e ispezionabile i progressi ottenuti in materia di uguaglianza di genere. Sin dal 1995 con la piattaforma di Pechino e le conferenze mondiali sulle donne convocate dall'ONU per promuovere l'uguaglianza tra i sessi, le statistiche di genere sono state considerate di fondamentale importanza per la pianificazione dello sviluppo, l'eliminazione degli stereotipi, la formulazione di azioni politiche ecc. Secondo la teoria del ruolo sociale, gli stereotipi di genere de-

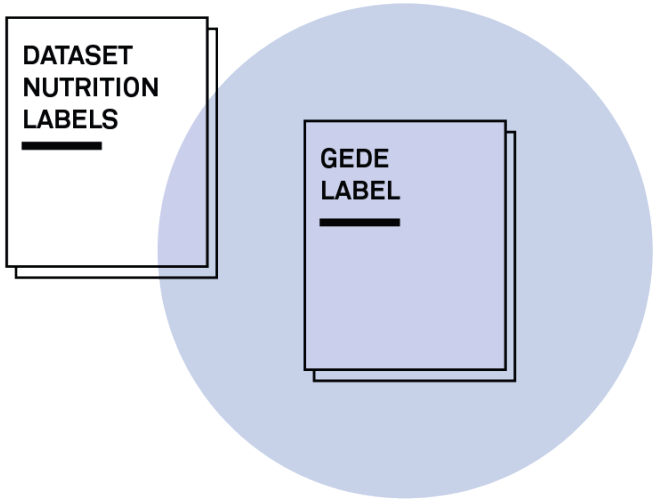


FIG. 46 Intersezione di GEDE con i fogli illustrativi.

rivano dalla distribuzione discrepante di uomini e donne in ruoli sociali sia in casa che sul lavoro (cfr. Eagly 1987; Koenig e Eagly 2014). Per questo motivo la valutazione di un dataset che utilizza variabili di genere non può essere esente da una verifica sull'equità.

Il processo per giungere a GEDE è concepito attraverso tre fasi principali, cui segue la sistematizzazione del problema che consiste nella messa a punto delle variabili presenti nella label, parallelamente ad una analisi grafica visiva riguardante l'esperienza dell'utente.

La prima fase si basa sulla ricerca, in particolare approfondendo il tema dell'IA, dei gender bias nell'IA ed infine delle possibili soluzioni esistenti per risolvere o quantomeno contenere i gender bias. Questa ricerca mira a creare un *background* di informazioni utili ad affrontare le fasi successive, fornendo gli strumenti per comprendere al meglio il senso dello stereotipo, come l'IA viene a contatto con i bias umani e come i ricercatori hanno finora indagato e tentato di contrastare il problema.

La seconda fase consiste nello studio dei sistemi di analisi dei dataset precedentemente teorizzati, parallelamente ad un vaglio delle variabili presenti nei dataset, dei metodi di analisi delle immagini riscontrabili in lettera-

tura, della concretizzazione degli stereotipi di genere negli artefatti visivi.

La terza ed ultima fase consiste nella vera e propria analisi, cui segue la strutturazione di GEDE. L'analisi, supportata dalle informazioni raccolte e dall'analisi dei casi studio, è determinante per comprendere quali informazioni è necessario includere nell'etichetta e per chiarificare i criteri di valutazione. La definizione delle scelte grafiche e visive segue questo passaggio e concretizza l'etichetta definendone scopi, target, formato e tipologia di *medium*.

GEDE evidenzia sia i punti di forza sia i punti di debolezza del dataset analizzato, chiarendone il contenuto in termini quantitativi e qualitativi, per rendere possibile una selezione più corretta e coerente in base all'uso che verrà fatto del dataset stesso.

Le strategie di analisi si basano sullo studio di altri lavori con presupposti comuni, quali le ricerche di Irene Biemmi svolte nell'ambito del progetto europeo *Polite*. Sia Biemmi che Pace affermano l'importanza di un doppio livello di analisi delle immagini, che per questo si suddivide in indagine quantitativa e successivamente qualitativa. La prima richiama le statistiche di genere e permette di calcolare la presenza dei due sessi, gli spazi in cui si collocano, le relazioni che intraprendono. Questo rappresenta il primo passo per capire se il dataset fornisce una rappresentazione equilibrata dei due sessi: uno squilibrio numerico è, infatti, «di per sé sintomatico di un atteggiamento discriminante» (Biemmi 2017). La seconda consiste nell'analisi qualitativa, per analizzare in profondità la parte visiva del dataset consistente nel vocabolario di immagini e fotografie presenti. Analizzare le immagini permette di indagare quali sono i ruoli assegnati ai generi dal dataset e come essi vengono rappresentati.

Per analizzare i dataset da queste prospettive occorrono due strumenti di indagine ad hoc: il primo step necessita di una scala che quantifichi la presenza di gender bias mentre il secondo è efficacemente narrato da una griglia di immagini utile all'analisi e comparazione qualitativa.

5.2 Selezione del dataset e del campionamento

Lo scopo ultimo è analizzare come gli algoritmi propongano (o ri-propongano) antichi stereotipi di genere nei dataset di immagini. La ricerca si muove su due fronti paralleli: da un lato l'analisi delle azioni e dall'altro l'analisi degli oggetti. Il senso di distinguere l'analisi e di non soffermarsi solo su un tipo di riconoscimento è utile per garantire una visione a tutto tondo sui ruoli di genere e per stimare l'assenza o eccessiva presenza sistematica di immagini di donne in particolari arene sociali. Gli oggetti e le azioni sono interrelati tra loro e, proprio per questo motivo, solo analizzando due dataset diversi si potrà avere uno spunto di partenza sufficiente per eventuali successive indagini. I due testi in esame, infatti, si propongono di essere i primi su cui strutturare gli esempi di fogli illustrativi. Non corrispondono al limite della ricerca sul gender bias nelle immagini dei dataset, ma ne vogliono rappresentare solo il principio.

Per trovare i testi da esaminare è stata necessaria una ricerca online, con lo scopo di recuperare due dataset semplici da utilizzare e da scaricare, con un'interfaccia intuitiva e facilmente raggiungibili dall'indicizzazione nei principali motori di ricerca. Sono stati quindi selezionati un dataset ampiamente analizzato dalla letteratura informatica per l'etichettatura visiva dei ruoli semantici *imSitu* (Yatskar et al., 2016, 2017); mentre per la classificazione multilaterale degli oggetti *Open Images v6* (Benenson et al., 2019), dataset noto e di uso comune. L'analisi ha il fine di confermare o meno la presenza di distorsioni a livello di gender bias in verbi ed oggetti e di stimare in quale quantità è eventualmente presente o assente. I testi in esame si diversificano principalmente per la composizione delle immagini e per l'ultima data dell'ultimo aggiornamento: se infatti, per il primo risale al 2018 per il secondo al più recente 2021. La scelta di due dataset con diversi range permette un confronto temporale, utile ad individuare eventuali discordanze o similitudini.

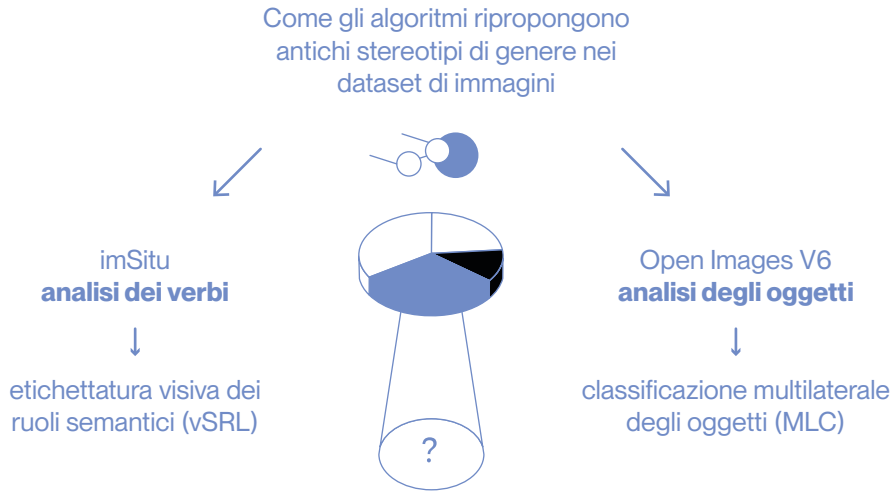


FIG. 47 Selezione dei dataset di ricerca.

5.2.1 imSitu

IS è un dataset per il riconoscimento delle situazioni, che agisce tramite l'etichettatura visiva dei ruoli semantici (vSRL). Come il sito ufficiale riporta, è un dataset che

FIG. 48 Tabella numeri di imSitu.

Totale immagini	126.102
Verbi	504
Situazioni per immagini	3
Totale annotazioni	1.481.851
Tipi di entità unici (>3)	1.788 (190)
Immagini per verbo (range)	250,2 (200-400)
Situazioni uniche (>3)	205.095

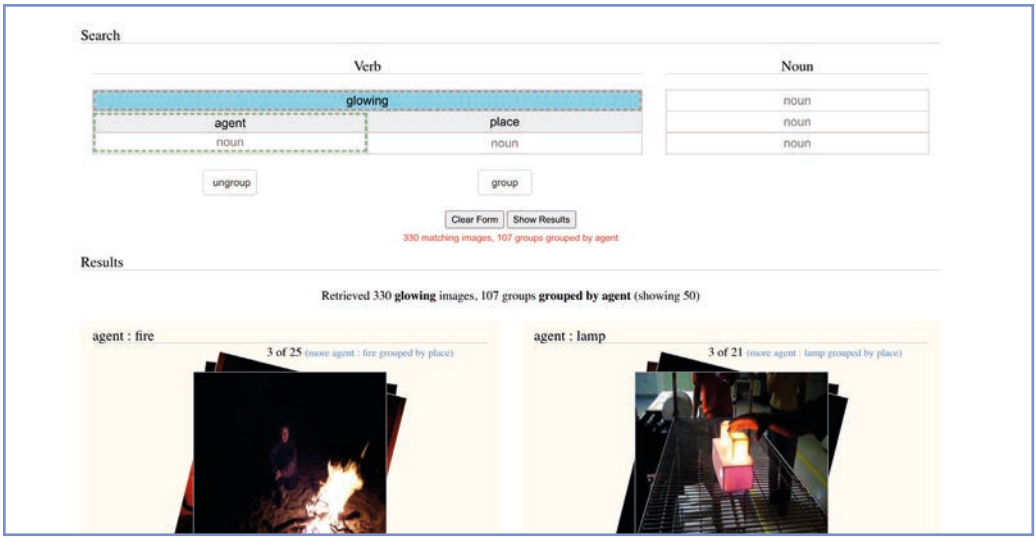


FIG. 49 Screen dal dataset online imSitu.

supporta il riconoscimento delle situazioni nelle immagini, e produce una sintesi della situazione rappresentata che include: l'attività principale, gli attori, gli oggetti, i luoghi e i ruoli che questi agenti svolgono nell'attività.

IS nasce nel 2016 presso l'University of Washington di Computer Science & Engineering e con il supporto di *Allen Institute for Artificial Intelligence* (AI2) e consiste in una collezione di 126.102 immagini che rappresentano 200.000 situazioni distinte e 504 verbi. I progettisti del dataset affermano che quest'ultimo possa essere usato per creare algoritmi robusti nel riconoscimento delle situazioni.

La parte verbale del set di azioni e ruoli di IS è derivata da FrameNet, mentre le entità visive da ImageNet. Le immagini sono state precedentemente raccolte da una ricerca su Google Images con tecniche di espansione delle *query*[2] e successivamente etichettate da Amazon Mechanical Turk[3]. Tutte le immagini sono state annotate da tre lavoratori a cottimo, che hanno selezionato una tra le 80.000 possibili *synset* di WordNet. I ricercatori affermano che questa etichettatura è affidabile in quanto 2 su 3 annotatori hanno fornito lo stesso synset per oltre il 75% dei ruoli (cfr. Yatskar, Zettlemoyer e Farhadi 2016).

2 Per aiutare i crowdworkers a capire come produrre le annotazioni, sono state generate delle etichette di esempio per ogni verbo. Cinque studenti hanno letto le definizioni di tutti i 1053 verbi e recuperato tre immagini corrispondenti a ciascun verbo da Google Image Search. Qualora non avessero trovato le immagini, il verbo sarebbe stato rimosso. Nel complesso, 580 verbi hanno superato questa fase di filtraggio (cfr. Yatskar et al. 2016).

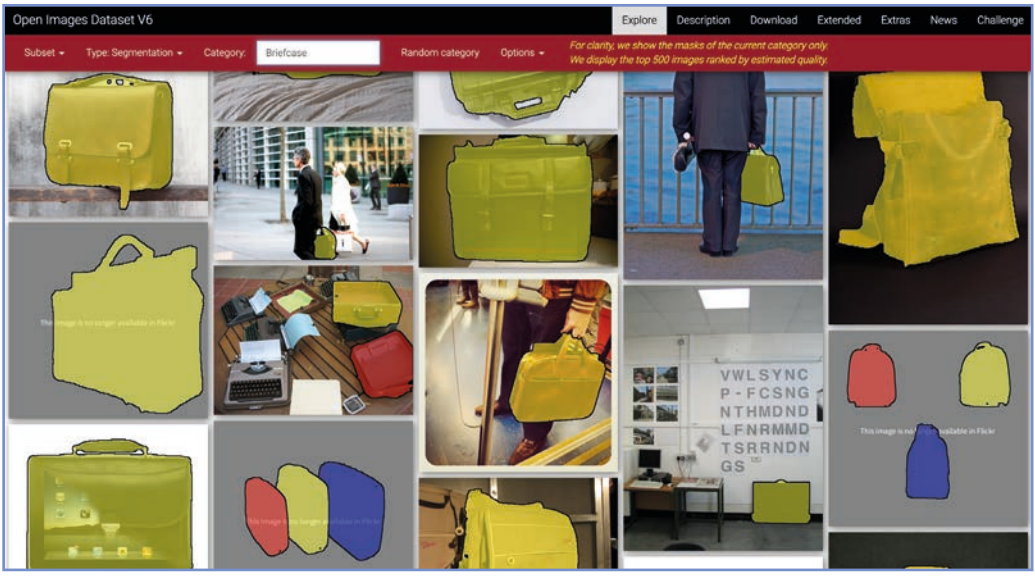
Tra le variabili del dataset IS si trovano:

- l'attività principale (ad esempio tagliare);
- gli attori coinvolti (uomo, pecore);
- gli oggetti usati (cesoie)
- il luogo di ambientazione (campagna);
- il ruolo che gli attori svolgono nell'attività (l'uomo sta tosando).

Nel paper di presentazione del dataset gli autori affermano che: «we also introduce structured prediction baselines and show that, in activity-centric images, situation-driven prediction of objects and activities outperforms independent object and activity recognition». Questo risulta essere un problema perché, come ha individuato la ricerca *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2017; paragrafo 4.4) il modello di ML addestrato su IS riconosce l'uomo in cucina come una donna, per il fatto che le immagini con il tag *cooking* comprendono un 33% di uomini a confronto con un 67% di donne. Questa percentuale è ancora più amplificata dopo l'allenamento: si riduce al 16% la presenza di uomini nel tag *cooking* (cfr. Zhao et al. 2017).

5.2.2 Open Images V6

Totale immagini	~9 milioni
Classi di oggetti	600
Oggetti per immagine (range)	8,3
Bounding boxes - annotazioni	16 milioni
Immagini di oggetti	1,9 milioni
Etichette a livello di immagine	59,9 milioni



OI V6 è un dataset per la classificazione multilaterale degli oggetti (MLC). Consiste in un dataset di circa 9 milioni di immagini annotate con etichette, box di delimitazione degli oggetti, maschere di segmentazione degli oggetti, relazioni visive e narrazioni localizzate. Le immagini sono molto

FIG. 50 Tabella numeri di OI V6.

FIG. 51 Screen dal dataset online OpenImages V6

diverse tra loro e spesso contengono scene complesse con vari oggetti (8,3 in media per immagine).

OI è un dataset per l'addestramento di sistemi di ML facilmente scaricabile ed utilizzabile da chiunque. Non richiede registrazione e nemmeno riferimenti personali per accedervi. La sua notorietà deriva anche dalla società che l'ha creato: Google; a questa sono sempre state associate grandi aspettative in termini di qualità e fairness.

Le classi sono identificate da MID (ID generati sistematicamente) che si possono trovare in Freebase o in Google Knowledge Graph API. Le annotazioni hanno licenza CC BY 4.0[3], rilasciata da Google LLC, mentre le immagini hanno una licenza CC BY 2.0.

> Note: while we tried to identify images that are licensed under a Creative Commons Attribution license, we make no representations or warranties regarding the license status of each image and you should verify the license for each image yourself.

Il paper di presentazione garantisce che il set di convalida e di test, così come parte del set di allenamento, hanno etichette, a livello di immagine, verificate dall'uomo. La maggior parte delle verifiche sono state eseguita da annotatori interni a Google. Una parte minore è stata eseguita con il crowd-sourcing di Image Labeler: *Crowdsource app*, g.co/imagelabeler. Sono stati utilizzati molteplici modelli di computer vision per generare i campioni delle classi, motivo per cui il vocabolario è notevolmente ampio. Come risultato del processo di annotazione, ogni immagine è annotata sia con etichette positive verificate a livello di immagine che indicano la presenza di alcune classi di oggetti, sia con etichette negative che indicano invece l'assenza di altre classi. Le etichette negative verificate sono affidabili e possono essere utilizzate durante l'addestramento e la valutazione dei classificatori di immagini. Nel complesso, ci sono 19.958 classi distinte con etichette a livello di immagine.

Tra le variabili del dataset OI si trovano:

» gli oggetti (le differenti classi);

3 Attribuzione 4.0 Internazionale (CC BY 4.0): indica la libertà di condividere -riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato - e modificare - remixare, trasformare il materiale e basarti su di esso per le tue opere per qualsiasi fine, anche commerciale - alla condizione di attribuzione. Devi riconoscere una menzione di paternità adeguata, fornire un link alla licenza e indicare se sono state effettuate delle modifiche. Puoi fare ciò in qualsiasi maniera ragionevole possibile, ma non con modalità tali da suggerire che il licenziante avalli te o il tuo utilizzo del materiale. (Definizioni di Creative Commons, consultabili al link https://creativecommons.org/licenses/by/4.0/deed.it)

» i box di delimitazione degli oggetti (i contorni degli oggetti);
» le maschere di segmentazione degli oggetti;
» le relazioni visive e narrazioni localizzate.

5.2.3 I parametri di campionamento

Il campionamento è diverso per la fase quantitativa e qualitativa: mentre la prima indaga fattori numerici riguardanti l'intero contenuto del dataset contenente le variabili di interesse, la seconda seleziona un piccolo campione dalle variabili selezionate nel primo step.

IS contiene molte attività non propriamente umane (ad esempio *rearing*, *retrieving* e *wagging*), azioni di animali, azioni meteorologiche, o di gruppi di persone che consentono di individuare i singoli (ad esempio *congregating*). Questi verbi sono stati filtrati, ottenendo un totale di 78 verbi e circa 11.700 immagini rispetto alle originali 120.000 nel dataset.

Anche in OI sono state filtrate tutte le categorie di oggetti non propriamente associate con esseri umani, rimuovendo gli oggetti che non presentano figure di sesso maschile o femminile almeno 100 volte nel dataset di training, lasciando un totale di 70 oggetti (il totale delle immagini non è calcolabile in quanto molto variabile a seconda dell'oggetto).

In entrambi i dataset sono state eliminate immagini dove il soggetto non fosse visibile, o avesse parti del corpo non riconoscibili o visibili per meno del 50%, immagini non inerenti al contesto come foto di animali, ritratti non umani e bambini piccoli nei quali il sesso non risulta comprensibile.

Per l'analisi qualitativa sono state selezionate le prime 50 immagini per verbo/oggetto che corrispondessero ai parametri sopra citati.

5.3 I criteri di analisi

Carta di identità: domande di inquadramento generale del dataset

Analisi dati quantitativi: dati numerici che rappresentano i numeri di genere nel dataset

Analisi dati qualitativi: analisi visiva semantica delle immagini contenute nel dataset

Campionamento dell'intero dataset

Campionamento basato sui tag di genere contenuti nel dataset

Dai tag di genere è scelto un campione di immagini osservabili (50 per verbo/oggetto)

FIG. 52 Sitemap di GEDE.

FIG. 53 (Pagina destra) Tipologia di domande principali che compongono i fogli illustrativi, suddivisi per fase di pipeline di ML.

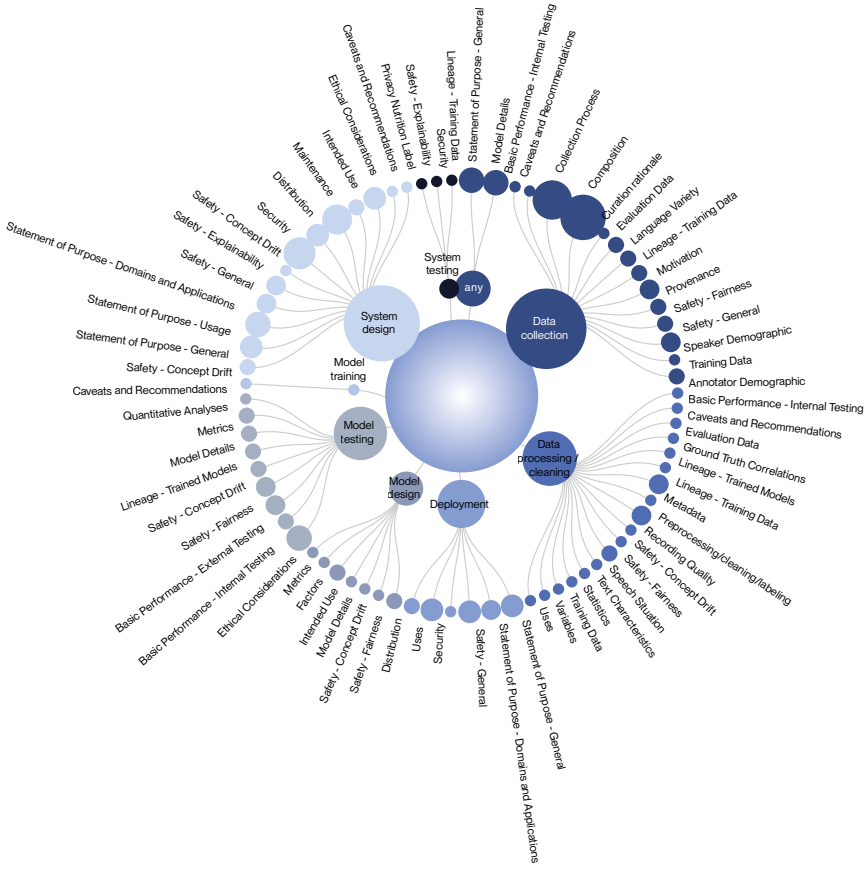
GEDE si compone di tre parti principali, con focus diversi per presentare il dataset in modo trasparente. Le tre parti sono rappresentate nel *wireframe* - scheletro - della label, individuate come:

1. carta di identità: fornisce una presentazione e specifica il contenuto del dataset, insieme alle informazioni di base, ad usi consigliati, ad eventuali avvisi presenti;
2. risultato dell'analisi quantitativa;
3. risultato dell'analisi qualitativa.

Nei successivi paragrafi sarà descritta nel dettaglio la metodologia di analisi per ognuna delle tre fasi.

5.3.1 Carta di identità del dataset

La prima parte di GEDE è caratterizzata dalla presentazione, spiegazione e identificazione del dataset in que-



stione. È una vera e propria carta di identità del dataset, derivata dalla letteratura.

Per definire questa prima parte si è reso necessario approfondire quali domande principali venissero poste al dataset dagli altri sistemi di fogli illustrativi esistenti e, successivamente, capire quali di queste domande fosse utile riportare. *About ML*[4] rappresenta per questo un riferimento e una risorsa fondamentale, che elenca le principali domande poste dagli altri dataset esistenti. Il documento di riferimento "Questions from existing research" di ABOUT ML suddivide le domande per argomento, per fase del ciclo di vita di ML, per la risposta attesa, per la fonte del paper di ricerca. Di seguito sono riportate alcune delle domande principali, successivamente incluse nella presentazione del dataset.

4 Si tratta di un'iniziativa pluriennale e multi-stakeholder guidata da PAI (Partnership on AI). Questa iniziativa, processo continuo e iterativo, mira a riunire una gamma diversificata di prospettive per sviluppare, testare e implementare pratiche di documentazione dei sistemi di ML in scala. Fonte: https://www.partnershiponai.org/about-ml/.

Domanda	Fase del ciclo di vita di ML	Risposta attesa
Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?	Composition - Data Collection	Long
How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.	Composition - Data Collection	Long
Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.	Composition - Data Collection	Long
Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.	Composition - Data Collection	Long
Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.	Composition - Data Collection	Long
Does the dataset relate to people?	Composition - Data Collection	Short

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.	Motivation - Data Collection	Long
Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?	Motivation - Data Collection	Short
Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.	Motivation - Data Collection	Short
List applications or scenarios for which the service is not suitable.	Security - Deployment	Long
Are you aware of possible examples of bias, ethical issues, or other safety risks as a result of using the service?	Safety - General - Deployment	Long
What person or organization developed the model?	Model Details - Any	Short
Is / Are your {contact information, cookies, demographic information, financial information, health information, preferences, purchasing information, SSN or Gov ID, activity on this site, location} used in {providing service and maintaining site, R&D, marketing, telemarketing, profiling, sharing with other companies, sharing with public forums}.	Privacy Nutrition Label - Data collection and system design	Long
Were there any relevant groups that were not represented in the evaluation dataset?	Caveats and Recommendations - Data Collection	Short
Is the model intended to inform decisions about matters central to human life or flourishing – e.g., health or safety? Or could it be used in such a way?	Ethical Considerations - System Design	Long

Le variabili considerate, dopo un'accurata selezione dalla letteratura e dalla casistica, sono:

1 Who:
   - » Chi ha creato il dataset (quale team, gruppo di ricerca) e per conto di quale entità (azienda, istituzione, organizzazione)?
   - » Chi ha finanziato la creazione del dataset? Se c'è una sovvenzione associata, si prega di fornire il nome del finanziatore e il nome e il numero della sovvenzione.
   - » Quale persona o organizzazione ha sviluppato il modello?

2 *About*: breve spiegazione del dataset
   - » Di cosa si tratta;
   - » A quale scopo è stato creato il dataset? Deve eseguire un compito specifico? Deve colmare una lacuna specifica? Si prega di fornire una descrizione.

3 Privacy: Open data / Safeguarded data / Controlled data[5].

4 Use Cases:
   - » Elencate le applicazioni o gli scenari per i quali il servizio è adatto;
   - » Elencate le applicazioni o gli scenari per i quali il servizio non è adatto;
   - » Ci sono i possibili usi al di fuori del campo di applicazione? Se si, quali?

5 Date:
   - » Quando è stato creato il dataset?
   - » Qual è la data dell'ultimo aggiornamento?

6 Data collection:
   - » Chi è stato coinvolto nel processo di raccolta dei dati (ad esempio, studenti, crowd workers, appaltatori) e come sono stati retribuiti (ad esempio, quanto sono stati pagati i crowd workers)?

FIG. 54 Esempi di domande principali che compongono i fogli illustrativi, suddivisi per fase di pipeline di ML.

5 La descrizione dei tre tipi di dati è stata recuperata dall'articolo Types of data access della testata online UK Data Service (link: https://ukdataservice.ac.uk/help/access-policy/types-of-data-access/), secondo la quale:
- *Open data*: sono dati concessi per l'uso con una licenza open, solitamente consistono in dati non personali e che hanno relativamente poche restrizioni d'uso.
- *Safeguarded data*: consistono in dati protetti, che possono avere condizioni aggiuntive come: accordi speciali; autorizzazione del depositante; uso limitato agli utenti non commerciali o accademici; clausole di distruzione dei dati; forme specifiche di citazione.
- *Controlled (o secure) data*: dati troppo riservati o sensibili per essere rilasciati tramite download.

» Come sono stati acquisiti i dati associati a ciascuna istanza? I dati erano direttamente osservabili (ad esempio, testo grezzo, valutazioni dei film), riportati dai soggetti (ad esempio, risposte ai sondaggi), o indirettamente dedotti/derivati da altri dati (ad esempio, tag part-of-speech, ipotesi basate su modelli per età o lingua)? Se i dati sono stati riportati dai soggetti o indirettamente dedotti/derivati da altri dati, i dati sono stati convalidati/verificati? Se sì, si descriva come.

7 Type of content: tipo di file ad esempio Tabular; csv; JSON.

8 Source: link di origine dataset.

9 Badges:
- » revisione della qualità - assente o presente;
- » informazioni sugli esseri umani - assenti o presenti;
- » revisione etica - assente o presente;
- » frequenza di aggiornamento dei dati - giornaliera, mensile, annuale, n.d.;
- » tipologia di fonte di finanziamento - no profit, per profit, governo, molteplice;
- » fonte dei dati - unica o molteplice;
- » presenta o non presenta dati sensibili;
- » informazioni sulle sottopopolazioni - presenti o assenti.

10 Alerts:
- » Eventuali possibili casi di bias, problemi etici o altri rischi per la sicurezza derivanti dall'uso del servizio offerto dal dataset.
- » è possibile che ci siano gruppi rilevanti che potrebbero non essere stati rappresentati nel dataset di valutazione?
- » altro

5.3.2 I dati quantitativi

La sezione di analisi qualitativa consente di valutare la sotto o sovra-rappresentazione di genere per certe azioni ed arene sociali. Ferree e Hall (1990) lo definiscono il "primo livello di rappresentazione" del bias nelle immagini. Dopo la scelta del campionamento di verbi e di oggetti, i dati vengono raccolti dal file sorgente in un file excel per il calcolo matematico di statistica.

Il file csv permette di individuare i principali tag legati al sesso utilizzati nel dataset per poi farne dato principale di riferimento. Di questi si valuta:

1. la percentuale di immagini con tag maschili e tag femminili sul totale, con il fine di esaminare se certi tipi di verbi/oggetti sono rappresentati in modo numericamente sbilanciato verso un sesso. In parallelo, una volta individuati i principali tag che identificano individui di sesso maschile, femminile e di sesso generico, risulta anche utile valutare il numero di immagini ad essi associate. Ad esempio nel caso di OI ed IS i tag individuati come più rilevanti per la loro ricorrenza sono: #Man, #Woman, #Girl, #Boy, #Person;

2. il secondo passaggio consiste nell'analizzare quali sono i verbi/oggetti in cui compaiono più tag ma-

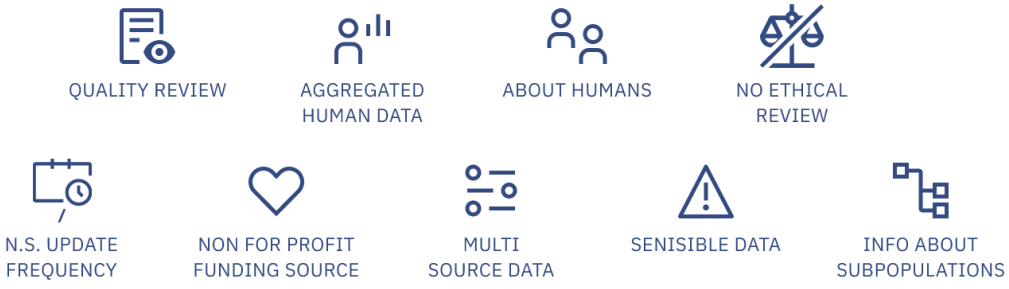


FIG. 55 Sistema di badges di GEDE.

schili o più tag femminili. Questo consentirà di individuare la panoramica dei verbi/oggetti le cui immagini sono principalmente composte da individui di un sesso, ricavandone il numero e confrontandolo in modo diretto con quello associato al sesso opposto;

3 la terza è un'analisi più approfondita del punto precedente: selezionando i primi 10 verbi/oggetti principali (ovvero quelli in cui sono presenti più tag), è necessario rendere chiara la percentuale di ricorrenza dei due sessi con il rispettivo verbo/oggetto.

I tre passaggi precedentemente individuati devono ora essere resi in forma grafica.

Partendo dal primo punto, il migliore dei modi per visualizzare i dati è utilizzare la tipologia di grafico denominata *100% stacked bar chart*, che analizza la relazione tra una singola componente e il totale in ciascun gruppo. Il grafico mette a comparazione la parte per il tutto.

Per la visualizzazione del secondo step (differenze tra tag che identificano individui di sesso maschile e femminile) il metodo più efficace consiste nel dividere i due grafici ed analizzare quali verbi/oggetti sono più o meno ricorrenti, in ordine di importanza. Il modello visivo che meglio interpreta questa necessità è il *donut chart*: un grafico a ciambella è funzionalmente identico ad un grafico a torta, con l'eccezione di un centro vuoto che gli consente la capacità di supportare più statistiche contemporaneamente. I grafici a ciambella (come quelli a torta) possono fornire una comparazione della parte rispetto al tutto. Il grafico consente di avere, per entrambi i sessi, una panoramica dei verbi/oggetti presenti con la relativa rilevanza numerica.

Per la terza infografica si sono calcolate le percentuali. Il processo è il seguente: per ogni verbo/oggetto si calcola il numero di immagini con tag #man, con tag #woman ed il numero di immagini totali. A questo punto la percentuale di immagini #man e #woman sul totale. Vengono poi selezionati i primi 10 risultati per entrambi i sessi con una maggio-

re differenza in numero, ordinati per la percentuale decrescente. Il modello visivo in questo caso consiste nel *Butterfly chart*, che consente un confronto diretto del numero di verbi/oggetti associati al sesso. Un grafico a farfalla (chiamato anche grafico a tornado) è un tipo di bar chart, in cui due serie di dati vengono visualizzati una accanto all'altra. Il grafico consente di fornire una panoramica rapida della differenza tra due gruppi con gli stessi parametri.

5.3.3 I dati qualitativi

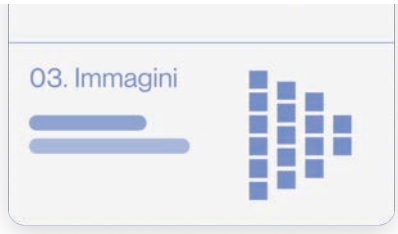


FIG. 56 Schema della metodologia progettuale di analisi.

A seguito del campionamento di verbi ed oggetti è necessario ridurre le immagini ad un numero osservabile: per quest'ultima operazione sono selezionate le prime 50 per verbo/oggetto. L'obiettivo del lavoro di ricerca qualitativa è valutare quanto le immagini del dataset contengano gender bias. C'è la necessità, come già anticipato, di ridurre al minimo l'interferenza di commenti e osservazioni soggettivi,

per questo è fondamentale ottenere parametri universalmente validi o quantomeno già dimostrati da studi precedenti. La scheda nasce dall'unione dello studio di pattern di stereotipi riscontrati in altri dataset, dall'analisi dell'origine delle immagini presenti nel dataset e dallo studio in letteratura dei diversi tipi di analisi visuali in ambiti paralleli.

La prima fase di esplorazione si concretizza nella raccolta di immagini dai due dataset. Da qui si dipartono due percorsi di indagine: il primo usa le immagini per individuare pattern di ricorrenze tra la dimensione di genere e i verbi o gli oggetti, ed il secondo esamina le immagini associate al genere dell'individuo presente in base alla schedatura progettata. Il processo è iterativo: in base ai nuovi pattern riscontrati la scheda può essere corretta ed aggiornata nel processo.

«Cosa vediamo quando vediamo qualcosa, dipende da che cosa è il qualcosa che vediamo. Ma il come vediamo quel qualcosa dipende da cosa sappiamo su quel qualcosa» (Falcinelli 2011)[6], indica l'esigenza di compiere un'analisi della soggettività e di minimizzare gli effetti sulla lettura dei risultati.

La schedatura analizza la presenza di elementi stereotipici nelle immagini. Lo stereotipo sessista si rileva quando ruoli, caratteristiche fisiche, caratteristiche psicologiche e caratteriali vengono attribuite in modo diverso in base ai due sessi uomo e donna, per questo motivo la scheda si suddivide nell'analisi dei due differenti sessi individuati dai tag del dataset.

6 Zenon Pylushyn Citato da Falcinelli in Guardare, pensare, progettare. Neuroscienze per il design, p. 161.

Categoria	Variabili figure tag femminili	Variabili figure tag maschili
Fisico (abbigliamento, sembianze)	▪ Presenza di make up ▪ Body Display* ▪ Corpo snello e attraente ▪ Sensualità evidenziata ▪ Figure esotiche**	▪ Body Display ▪ Bianco, corpo snello, atletico ▪ Giovane e attraente ▪ Vestito in modo professionale
Palette cromatica (nell'abbigliamento)	▪ Tonalità accese e varie* ▪ Impiego del rosa	▪ Tonalità scure o neutre ▪ Impiego dell'azzurro o blu
Gesti	▪ Messa in posa stereotipata ▪ Feminine touch* ▪ Emotività esagerata: sorriso, trasecolata, sorpresa, ecc. ▪ Licensed withdrawal** ▪ Sistemare l'abbigliamento, il makeup o i capelli	▪ Messa in posa stereotipata ▪ Impegnati nell'azione ▪ Mostrare i muscoli ▪ Sistemarsi i capelli ▪ Emotività: assente o comunica serietà
Ruoli	▪ Function Ranking* ▪ Non professionali ▪ Sessualizzate ▪ Ruolo passivo ▪ Dipendenza fig. maschili ▪ Donna come madre ▪ Lavoro di cura**, domestico	▪ Function Ranking ▪ Professionali ▪ Ruolo attivo, autoritario ▪ Indipendenza da fig. femminili (se presenti) ▪ Mostrare forza ▪ Ostentare ricchezza
Luoghi	▪ Ambienti interni e casalinghi ▪ Shopping, beni di consumo ▪ Luoghi di lavoro: esempio scuola, campo estetico* ▪ Presenza di fiori	▪ Ambiente esterno, campo aperto ▪ Luoghi di lavoro: esempio officina, costruzioni, ecc.
Inquadrature	▪ Bodyism* ▪ Posizione subordinata ▪ Dimensione relativa*** ▪ Peso percettivo minore ▪ Figure più giovani ▪ Posizioni chine o sdraiate. ▪ Posteriore alle fig. maschili	▪ Faceism** ▪ Posizione dominante ▪ Dimensione relativa ▪ Peso percettivo maggiore ▪ Figure più anziane ▪ In piedi con capo eretto. ▪ Anteriore alle fig. femminili

	Specifiche	Fonte
	*Body display si intente alto grado di nudità: le figure sono svestite o indossano vestiti non coprenti. **Figure esotiche si intende figure femminili presentate in contesti particolarmente strani o surreali.	Kang 1997; Lukas 2002; Halliwell e Dittmar 2004; Paoli 2008; Scanu 2012; Lower 2018; Planned Parenthood 2021
	*Tonalità accese e varie: colori pastello, cromie vivaci e molto diverse tra loro.	Giomi 2013; Abbatecola e Stagi 2017
	*Feminine touch: tendenza a ritrarre figure femminili che toccano parti del proprio corpo, cullano e/o accarezzano oggetti. **Licensed withdrawal: figure femminili impegnate in coinvolgimenti che le rimuovono psicologicamente dalla situazione sociale, lasciandole disorientate e dipendenti.	Goffman 1979; Torrioni 2014; Bucchetti 2020
	*Function Ranking: se la figura maschile e femminile collaborano la prima ha ruolo esecutivo. **Lavoro di cura: se la figura femminile si trova in presenza di bambini e/o anziani .	Goffman 1979; Courtney e Lockeretz 1979; Berger 2008; Eisend 2010; Abele e Wojciszke 2014
	*Lavoro nel campo estetico: ad esempio modella, vestiti, spettacolo ecc.	Eisend 2010; Scanu 2012; Abbatecola e Stagi 2017; Planned Parenthood 2021
	*Bodyism: focus su parti del corpo. **Faceism: focus sul volto. ***Dimensione relativa: la figura femminile risulta più bassa della figura maschile.	Mulvey 1975; Goffman 1979; Courtney e Lockeretz 1979; Wood 1994; Kang 1997; Halliwell e Dittmar 2004; Eisend 2010; Scanu 2012; Caratti 2015; Bucchetti 2017; Abbatecola e Stagi 2017

FIG. 57 Scheda di valutazione del livello di bias nelle singole immagini.

01. Apparenza fisica

Le variabili per l'analisi dell'apparenza fisica derivano da diversi studi, tra i quali le considerazioni di Scanu (2012) che si soffermano sulla tipologia di fisico (magro, snello, formoso, atletico/muscoloso). Ci si aspetta dalle donne che siano magre e aggraziate o, per dirlo con le parole di Halliwell e Dittmar (2004) «women of a slim body type», mentre dagli uomini che siano alti e muscolosi «men of an athletic body type». Ci si aspetta anche che uomini e donne si vestano e si curino in modi che sono stereotipati per il loro genere: gli uomini indossano pantaloni e acconciature corte, le donne indossano abiti e trucco (cfr. Planned Parenthood 2021). L'abbigliamento è uno dei principali strumenti di comunicazione non verbale, e in questo si caratterizza sia per le proprietà cromatiche sia per le forme. Citando Paoli (2008): «clothing presents the wearer with a choice of images in the sense that the difference between clothing items is "not just one of fabric and style, but one of identity" (Williamson 39)».

Lo studio di Lower (2018), *Style Speaks: Clothing Judgments, Gender Stereotypes, and Expectancy Violations of Professional Women* mostra la relazione tra i capi indossati e le qualità attribuite alla persona d'istinto. Cita l'esempio di una donna, vista come non esperta se sul posto di lavoro indossa abiti femminili, invece tendenzialmente considerata come severa se indossa abiti maschili, ad esempio un completo elegante.

Kang (1997) riprende lo studio di Goffman sulle immagini pubblicitarie ed aggiunge alle variabili il "Body Display" inteso come l'alto grado di nudità. Le figure femminili sono ritratte con vestiti che rivelano il corpo, tra i quali minigonne, gonne strette o abiti da sera che espongono la scollatura, pantaloncini corti, vestiti "see-through", costumi da bagno. La nudità comprende i modelli traslucidi sotto l'abbigliamento e la lingerie, modelli vestiti con un asciugamano, o modelle ritratte senza vestiti. Le riprese "ravvicinate" in cui le spalle dei modelli sono nude sono considerate nudità.

Un ulteriore parametro è la rappresentazione di figure esotiche, ossia donne in contesti incredibilmente strani

o surreali: con animali, situazioni bizzarre, creature mitiche, o altri scenari irreali o estremizzati (cfr. Lukas 2002).

Le variabili individuate si suddividono in:

» figure di sesso apparentemente femminile: giovane, corpo snello; sensualità evidenziata; presenza di make up evidente; body display: alto grado di nudità (presenza di scollature, trasparenze, abbigliamento intimo in evidenza, minigonne, gonne strette, abiti da sera, pantaloncini corti, vestiti "see-through", costumi da bagno, lingerie, asciugamano o assenza di vestiti); abbigliamento informale; figure esotiche;

» figure di sesso apparentemente maschile: pelle bianca, corpo atletico/muscoloso; giovane e attraente; body display: alto grado di nudità; abbigliamento elegante/formale; pantaloni.

02. Colori, cromie (abbigliamento)

L'elemento cromatico degli abiti indossati è associato all'occupazione o all'azione svolta dall'individuo, mentre in altre occasioni diventa un carattere di demarcazione, utilizzato sin dalla nascita per segnare i confini tra maschile e femminile. I demarcatori cromatici caratterizzano soprattutto il mondo femminile, dato che il mondo maschile assume sempre un carattere di universalità. Si veda Abbatecola:

> Il blu potrà essere indossato anche da una bambina senza che questo rappresenti necessariamente una violazione dei confini, mentre il rosa, in quanto marcatore di una forma specifica che assume l'umanità, non potrà essere indossato da un bambino/ragazzino, senza che ciò venga percepito come attraversamento indebito che non può passare inosservato. (Abbatecola 2017: 86)

Il colore caratterizza anche il passaggio tra le diverse età. Giomi (2013a; 2013b) analizza le pubblicità per l'abbigliamento infantile, dove appaiono evidenti alcuni elementi che suddividono in modo chiaro il territorio maschile da quello femminile: i codici interpretativi riguardano il colore, l'ambientazione e la divisione di ruoli. Per i bambini prevalgo-

no colori forti e accesi, per le femmine i toni del rosa e colori tenui; mentre i bambini vengono ritratti in movimento e negli spazi aperti, privilegiando azione e fisicità, per le bambine, a sottolinearne la staticità, prevale lo spazio privato e la cura di sé e del prossimo.

Nell'età adulta lo scenario si modifica: al femminile sono associate una varietà di cromie accese, mentre al maschile sono associati i colori neutri.

> Pinkizzazione è la traduzione di "Pinkification", un neologismo inglese utilizzato per indicare il processo per cui un prodotto o un servizio si avvale del colore rosa per attrarre il pubblico femminile. Un "rinnovamento cromatico" che in alcuni casi comprende anche una caratterizzazione formale con linee morbide, curvilinee e una maggiore presenza di decorazioni (fiori, cuori e altri segni) ed effetti luccicanti (glitter, brillantini). L'associazione tra rosa e femminile, in particolare nel contesto occidentale, è molto radicata, sebbene non così lontana nel tempo. (Abbatecola e Stagi 2017)

I confinamenti delineati dalle cromie sono chiari, impossibili da discutere senza rischiare uno stigma. Riprendendo ancora Abbatecola e Stagi (2017), il rosa, i fiocchi e i lustrini sono i simboli degli attributi di leggerezza, delicatezza e frivolezza, qualità femminili che le bambine devono apprendere sin da piccole. Ai maschi invece è lasciata una gamma più vasta di colori e di territori. In presenza di una divisa professionale le differenze cromatiche scompaiono «per lasciare il posto ad un immaginario scuro e neutro, legato al maschile, utilizzato per raffigurare manager, politici, uomini d'affari»[7]. I colori non sono infatti fine a se stessi e agli oggetti che li presentano, ma sono direttamente correlati ai gusti, al carattere e alla personalità del genere. Se l'uomo è considerato l'universale umano − non solo dagli uomini ma anche dalle stesse donne − va da sé che l'azzurro e il blu, benché identificativi della maschilità, possano essere percepiti come marcatori identitari più deboli rispetto al rosa. Se il maschile è percepito come non caratterizzato dal punto di vista del genere, il blu potrà essere indossato anche da una bambina senza che questo rappresenti necessariamente una violazione dei confini, mentre il rosa, in quanto marcatore di una

7 Oswald P.A., *Sex-Typing and Prestige Ratings of Occupations as Indices of Occupational Stereotypes* (2003: 289-296).

forma specifica che assume l'umanità, non potrà essere indossato da un bambino/ragazzino, senza che ciò venga percepito come attraversamento indebito che non può passare inosservato (cfr. Abbatecola e Stagi 2017).

Lo scopo di questo punto è la verifica dell'utilizzo delle cromie e della tonalità dominante in relazione al sesso dell'individuo:

- in figure femminili si verificherà l'impiego di tonalità accese e varie o del rosa;
- in figure maschili l'impiego di tonalità scure, neutre o dell'azzurro/blu.

03. Gesti (messa in posa; azioni fisiche)

I gesti sono la messa in pratica del linguaggio di genere, di una comunicazione non verbale che spesso nelle immagini si appropria di concetti stereotipici di genere.

Non di rado (vedi paragrafo 4.4) le figure femminili sono ritratte in una messa in posa stereotipata (Bucchetti 2020), di cui ad esempio la donna trasecolata, sbalordita, imbarazzata o ridente.

Paola Maria Torrioni (2014: 45) sottolinea che «la figura maschile è idealizzata attraverso le dimensioni della forza, della razionalità e indipendenza, mentre per quella femminile prevalgono tranquillità, dedizione alla cura, capacità di ascolto, dipendenza» e questo si rispecchia nei gesti che gli individui compiono. Mentre la donna ha poca capacità di controllare le emozioni quindi è mostrata come troppo felice e sorridente o isterica; l'uomo tende a non mostrare emozioni e quando le mostra tendenzialmente si associano a gesti di rabbia. Sebbene Plakoyiannaki, Mathioudaki, Dimitratos e Zotos (2008) sostengono che le donne siano «concerned with physical appearance», i gesti narcisisti di sistemarsi capelli, make up o vestiti appartengono ad entrambi i sessi.

Goffman utilizza due categorie con le quali le figure femminili sono narrate nelle immagini pubblicitarie delle riviste da lui analizzate, individuate da: "Feminine touch" e "Licensed Withdrawal".

Quindi, in base alla metodologia di analisi ed ai diversi pattern individuati mediante analisi sul campo, le variabili da ricercare si suddividono in:

» figure femminili: messa in posa stereotipata; feminine touch (toccare parti del proprio corpo, cullare e/o accarezzare oggetti); trasudare emotività (sorriso molto ampio, trasecolare, mostrare sorpresa arrabbiatura ecc.); licensed withdrawal, ossia essere rimosse dalla situazione (guardare altrove, mantenere conversazioni telefoniche); sistemarsi accessori, capi d'abbigliamento, makeup o capelli.
» figure maschili: messa in posa stereotipata; essere impegnati nell'azione; mostrare i muscoli; sistemarsi capelli, abiti; emotività assente o comunica serietà o rabbia.

04. Ruolo (occupazione o ruolo che ricopre)

Come già è stato analizzato nel paragrafo 4.4, la categoria "Function Ranking" (o classifica delle funzioni) teorizzata da Goffman rimane molto attuale nel sostenere che, se un uomo e una donna collaborano, è probabile che l'uomo svolga il ruolo esecutivo. Il medesimo concetto è sostenuto da Courtney e Lockeretz (1979) i quali specificano che, solitamente le donne non sono mostrate come professionali, al contrario della controparte maschile: «gli uomini agiscono e le donne appaiono» (Berger 2008). Anche Eisend traccia un confine netto in cui definisce «Men as an authority», a confronto con «Women as passive» (Eisend 2010).

Non di rado il concetto di occupazione è stereotipato:insegnanti e le infermiere sono donne e piloti, medici e ingegneri sono uomini (cfr. Planned Parenthood 2021). Il concetto è interrelato alle caratteristiche di *Agency* e di *Communality* (cfr. Abele e Wojciszke 2014). Per questo la presenza di bambini o di anziani è decisamente maggiore in numero nel caso di soggetti femminili, dove identifica il ruolo di cura. Le variabili individuate si suddividono in:

» figure apparentemente femminili: verificare se sussiste il "function ranking", ossia se figura maschile e femminile collaborano la prima ha

ruolo esecutivo; se risultano non professionali; di ruolo passivo; in dipendenza delle figure maschili (se presenti); se sono rappresentate come madri; se eseguono il lavoro di cura, nel caso di presenza di bambini/anziani; se eseguono lavori domestici.

» figure apparentemente maschili: verificare se sussiste il "function ranking": ossia se figura maschile e femminile collaborano la prima ha ruolo esecutivo; se hanno un atteggiamento professionale, serio; ruolo attivo, autoritario; Indipendenti da figure femminili (se presenti); mostrare forza; ostentare ricchezza.

05. Ambientazione (luoghi, oggetti)

Una notevole importanza nell'analisi delle immagini è affidata ai luoghi:

> Le differenze costruite dalla pubblicità nei ruoli/interessi/gusti attribuiti a maschi e a femmine sono riassunte da un'opposizione di fondo che riguarda in particolare gli ambienti. Il maschile è rappresentato in spazi aperti, associabili ai valori della scoperta, dell'esplorazione, dell'avventura, dell'interazione con il territorio, della sfera pubblica. Il femminile è quasi sempre inserito in spazi chiusi, nello spazio domestico, lo spazio delle relazioni personali, dell'intimità, una sfera privata per definizione separata dal mondo esterno. (Abbatecola e Stagi 2017)

Ci si aspetta che le donne si prendano cura dei bambini, cucinino e puliscano la casa, mentre gli uomini si occupano delle finanze, lavorano alle automobili e fanno le riparazioni in casa (cfr. Planned Parenthood 2021). Per questo si tende a raffigurare le figure femminili «in a domestic environment» (Eisend 2010).

Anche gli oggetti utilizzati o presenti nella scena contribuiscono a fornire il senso dell'ambientazione. Nella loro ricerca, Abbatecola e Stagi, affidano ai bambini il compito di raffigurare su un foglio diviso a metà gli oggetti che "le femmine hanno, i maschi hanno", con il fine di rintraccia-

re il processo di genderizzazione. Dal materiale raccolto «si evince che per le bambine prevale la cura di sé e l'apparire, mentre per i bambini l'azione lo spazio pubblico, l'elettronica» (Abbatecola e Stagi 2017).

> Nello studio di Cherney e London (2006) emerge nettamente questa divisione delle attività: computer, sport e televisione – quest'ultima soprattutto come supporto ai videogiochi ma anche per la visione di un certo tipo di cartoni animati – sono gli ambiti per esercitare e sviluppare alcune di competenze considerate maschili: un addestramento alle abilità spaziali, all'esplorazione, alla competizione, in parte anche all'aggressività (Abbatecola e Stagi 2017). Blaise ha chiamato identità girly girl l'insieme di pratiche per "indossare la femminilità" da parte delle bambine: gli atteggiamenti corporei, i movimenti di capelli o dei vestiti (twirling e curtseying), il trucco, il modo di parlare, sono elementi della messa in scena della femminilità (2005:85). Le pratiche incarnate dai bambini, invece, si sviluppano intorno alle competenze, la produzione di lavoro visibile e la valutazione dei successi Individuali. (*ibidem*; corsivo mio)

Le variabili individuate si suddividono in:

» figure apparentemente femminili: verifica di ambientazioni in interni e spazi casalinghi; in negozi in cui comperare beni di consumo, siano essi primari o secondari; luoghi di lavoro tipicamente *communal* quali scuola, campo estetico (modella, vestiti, spettacolo); presenza di fiori.

» figure apparentemente maschili: verifica di ambientazioni in esterni, raffigurati con un campo aperto; luoghi di lavoro tipicamente *agentic* quali officina, costruzioni, uffici ecc.

06. Inquadrature

La verifica del piano di rappresentazione, dell'inquadratura, dei pesi percettivi dell'organizzazione spaziale è una parte fondamentale per lo studio di genere nelle immagini. Claudia Scanu cerca di definire lo *sguardo maschile* nell'analisi delle immagini pubblicitarie nelle riviste, considerando l'angolazione da cui è stata scattata la fotografia (che

può essere dall'alto, dal basso o frontalmente), i piani di inquadratura (figura intera, piano americano, mezza figura, primo piano, primissimo piano o dettaglio), la profondità di campo (estesa o ridotta), la messa a fuoco (sul primo piano, sul piano medio o sullo sfondo) (cfr. Scanu 2012: 214). Il corpo femminile, secondo la concezione del *male gaze*, è un corpo da guardare, è oggettivizzato e aderisce a stretti canoni di bellezza, che variano in base ai periodi storici. Quello maschile, al contrario, è un corpo in azione, che domina. L'essere maschio è definito dalle prestazioni corporee (cfr. Connell 1996): come afferma Messner si misura l'adeguatezza della maschilità dei ragazzi rispetto al modello vincente in base alla riuscita o meno negli sport competitivi (cfr. Abbatecola e Stagi 2017). Quello maschile è un corpo che deve sempre dimostrare la virilità attraverso il controllo delle emozioni.

Per questo l'angolazione dall'alto è «utilizzata principalmente quando la donna è seduta/accovacciata o distesa, ponendola in una posizione di inferiorità rispetto all'osservatore, che domina così la scena» (cfr. Scanu 2012: 214). Il concetto deriva dalla categoria di analisi "Relative Size", o dimensione relativa in *Gender Advertisements* di Goffman, descritta come il posizionamento in primo o secondo piano o di ruoli più o meno centrali nella scena. Il posizionamento della figura può rendere chiara la loro dipendenza o indipendenza rispetto all'altra. Anche qui lo stereotipo distingue le figure in modo chiaro: tende a rappresentare le donne come dipendenti, incompetenti, come *caregiver* primari, come vittime e oggetti sessuali, a confronto di uomini indipendenti, autoritari, capofamiglia e aggressori (cfr. Wood 1994). Il concetto è descritto da Goffman nella categoria "Ritualization of Subordination", o rituale della subordinazione, dove posture inclinate possono essere lette come un'accettazione della subordinazione, un'espressione di ingraziazione, sottomissione e appagamento, che si contrappongono a posture erette di sfida, tipicamente maschili (cfr. Goffman 1979). Tutti questi metodi enfatizzano il peso percettivo minore delle figure femminili (cfr. Bucchetti 2017). Le variabili individuate si suddividono in:

» figure apparentemente femminili: l'obiettivo ha focus su parti del corpo (bodyism); la figura appare in posizione subordinata; la dimensio-

ne relativa risulta più bassa delle figure maschili; peso percettivo minore; figure più giovani; figure collocate in piedi o sedute più in basso rispetto alle figure maschili; posizioni chine o sdraiate; figure collocate posteriormente alle figure maschili.

» figure apparentemente maschili: l'obiettivo ha focus sul volto (faceism); le figure appaiono in posizione dominante; la dimensione relativa risulta più alta della figura femminile; peso percettivo maggiore; figure più anziane; figure collocate in piedi o sedute più in alto rispetto alle figure femminili; in piedi con capo eretto; collocate anteriormente alle figure femminili.

Punteggio

Il punto, o *bias value*, è attribuito all'immagine se anche solo una variabile per segno stereotipico risulta presente. Il punteggio non varia quindi se nella fotografia si verificano più variabili appartenenti allo stesso segno: l'immagine è valutata in base alla presenza o assenza dei segni presenti. Si noti quindi che la gravità dell'immagine non coincide con il punteggio, il cui scopo è quello di valutare la presenza/assenza di gender bias nelle immagini del dataset. La gravità di una immagine, inoltre, è un parametro soggettivo che non rientra nel modello di schedatura proposto.

Risulta necessario fornire un'indicazione del *bias value* contenuto nel dataset in media. È stato per questo utilizzato il metodo di valutazione a 6 stelle. Affidato ad ogni immagine un valore da 0 a 6, per ogni verbo/oggetto è stata calcolata una media ponderata, scoprendo il valore di bias medio per quel singolo verbo/oggetto. Per ricavare il valore di bias del dataset è bastato calcolare la media aritmetica del valore di bias di ciascun componente (verbo/oggetto).

Le infografiche di questa terza e ultima componente di GEDE sono di prevalenza visiva: l'elemento visuale è protagonista. L'infografica principale è quindi una scheda che si compone di tre parti: lista di verbi/oggetti e, per ognuno di questi, il bias value è rappresentato sotto forma di he-

atmap. Accanto, in ordine crescente, si trova un esempio di immagine, appartenente a quel verbo/oggetto, per ogni bias value. Per visualizzare tutte le foto del campionamento è possibile espandere la colonna selezionata, in modo da svelare un ulteriore dato: la quantità di fotografie per ogni bias value.

L'ultima parte dell'etichetta rappresenta uno zoom che esemplifica il gender bias in senso binario del termine. Vengono estratte dall'infografica precedente immagini simili che rendono chiara la differenza tra il bias delle rappresentazioni dei soggetti di sesso maschile e dei soggetti di sesso femminile. Le immagini sono state selezionate per somiglianza di ripresa, inquadratura, disposizione delle figure e colori presenti. Al momento le immagini di questa fase sono state scelte in modo manuale dal ricercatore, ma in un momento successivo potrebbe essere implementato un software di automatizzazione che individua le immagini più simili per tipi di caratteristiche definite. Questi tipi di IA prendono il nome di *visual similarity*.

5.4 GEDE: Gender Debiaser

GEDE si rivolge ad un pubblico informato sull'argomento bias nell'IA, sia l'utente un data scientist o un ricercatore. Per la progettazione è stato seguito l'Activity Centred Design (ACD): modello di progettazione che si concentra sulle attività finali che l'artefatto può svolgere, virando l'attenzione dal solo utente (centrale nell'User Centered Design) al sistema intero. L'ACD è una delle molte prospettive che si possono impiegare nella progettazione. Se l'esperienza sarà ben pensata, la maggior parte degli utenti saprà adattarsi e lo strumento verrà ritenuto utile e funzionale[8].

Il modello di ACD può essere ideale per l'implementazione di nuovi paradigmi o ripensamenti innovativi. La ricerca si è svolta quindi analizzando prima come intervengono gli altri tipi di fogli illustrativi, a quali domande rispondono e qual è l'interazione che si suppone l'utente dovrebbe avere. Immediatamente successiva è stata la strutturazione dell'analisi di immagini, dove anche qui si è intervenuti ricercando le principali fonti di immagini dei dataset, in modo da coprire tutti i punti possibili di osservazione.

Lo svantaggio principale di cui il progetto risente riguarda la mancanza di un confronto con un punto di vista esterno al processo, che risulterebbe fondamentale nella parte più scientifica e legata all'analisi numerica del dataset e dell'uso del dataset. GEDE è però sia un artefatto che un processo, nel senso che si può creare e co-creare con chi ha progettato i dataset che si andranno ad analizzare, interagendo in prima persona e correggendo i parametri che hanno bisogno di correzione.

SCENARI DI UTILIZZO

Il data scientist o chi deve decidere quale dataset selezionare per il nuovo modello si trova davanti ad una scelta non guidata: non è facile stabilire in breve tempo quale fra la moltitudine di corpus ritrovabili online faccia al caso del modello in questione. Di seguito sono riportati tre esempi di dataset problematici in cui l'utente si potrebbe imbattere:

8 Martin L., «User-centered Design (UCD) and Activity-centered Design (ACD)», link <sitemotif.com/2008/07/user-centered-design-ucd-and-activity-centereddesign-acd/> (29.04.18).

» Nel caso in cui il dataset presenti una sovra- o sotto rappresentazione di genere per certi campi il modello, come riporta il paper *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints* (Zhao et al. 2017), amplificherebbe la sovra o sotto rappresentazione. Nel caso sopra citato, accade che nel dataset di partenza il verbo "cooking" è associato alla maggior parte delle foto in cui compare un soggetto di aspetto femminile. Se il modello allenato su questo dataset vede una foto in cucina con un soggetto non ben visibile, conclude che il soggetto sia una donna. È una conclusione logica, considerato che nei dati da cui ha imparato il soggetto associato all'ambiente cucina è prevalentemente "woman". I ricercatori dimostrano che, dato un corpus gender biased, i modelli strutturati come il conditional random fields, amplificano il bias (Zhao et al. 2017).

One common theme is the notion of bias amplification, in which bias is not only learned, but amplified [4,7,6]. For example, in the image captioning scenario, if 70% of images with umbrellas include a woman and 30% include a man, at test time the model might amplify this bias to 85% and 15%. (Burns et al. 2019)

» Allo stesso modo, la raffigurazione di diversi gruppi sociali può essere appresa automaticamente da modelli di immagini non supervisionati. I ricercatori del progetto *REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets* hanno scoperto che, nel dataset OpenImages, una proporzione maggiore di immagini "femminili" sono ambientate in "casa o hotel" rispetto alle immagini "maschili". Quest'ultimo è più spesso raffigurato in scene "industriali e di costruzione". La differenza di rappresentazione potrebbe essere esemplificativa per spiegare i pregiudizi di genere incorporati nelle immagini di AI non supervisio-

nate. In generale, se la rappresentazione delle persone nelle immagini online riflette i pregiudizi sociali dell'uomo che sono documentati nella cognizione e nel linguaggio, Steed e Caliskan (2021) concludono che i modelli di immagini non supervisionati potrebbero imparare automaticamente pregiudizi simili a quelli umani da grandi collezioni di immagini online. Succede in *Unequal Representation and Gender Stereotypes in Image Search Results for Occupations*, (esempio già visto nel paragrafo 3.3.3) dove alle immagini di politici donne sono correlati *tag* legati all'estetica o che descrivono apparenze fisiche (*hairstyle*, *smile*), mentre alle immagini di politici uomini tag che descrivono l'occupazione e il livello di prestigio (ad esempio *businessman*).

» Anche i sistemi di predizione del "next-pixel" possono propagare stereotipi nell'utilizzo. Ad esempio l'applicazione incauta di un modello generativo come il *iGPT* può produrre descrizioni distorte delle persone. Steed e Caliskan (2021), come caso studio qualitativo, hanno selezionato 5 immagini di volti dall'aspetto maschile e 5 femminile da un database generato con *StyleGAN*. Hanno poi rifilato i ritratti sotto al collo e usato iGPT per generare 8 diversi completamenti. Il risultato è il seguente: per i volti femminili il 52,5% dei completamenti presentava un bikini o un top scollato; per i volti maschili il 7,5% dei completamenti era a torso nudo, mentre il 42,5% indossava completi o altro abbigliamento specifico di tipo lavorativo. Uno di questi aveva una pistola in mano. Il comportamento presentato dall'algoritmo potrebbe derivare dalla rappresentazione sessualizzata delle persone, specialmente delle donne, nelle immagini di Internet e serve a ricordare la storia controversa della computer vision con le immagini oggettivanti (cfr. Steed e Caliskan 2021).

Per questi motivi l'algoritmo addestrato su un dataset *biased* non risulta affidabile. Qui entra in gioco GEDE: la lettura dell'etichetta consentirà di effettuare una scelta di dataset informata e conscia del contenuto e di eventuali punti di forza e punti di debolezza del dataset. L'obiettivo finale è garantire al data scientist, o chi per lui, la scelta di un corpus di dati giusto per il giusto modello.

5.4.1 Generazione della GEDE label

Avendo la necessità di poter creare il foglio illustrativo in modo semplice ed immediato, senza causare perdite di tempo ai data scientist o ricercatori, dopo un confronto con sviluppatori si è optato per la strutturazione di un form. Il form è compilabile e sarà composto di domande a risposta chiusa, aperta o multipla che andranno a mappare tutte e tre le parti di GEDE. Al termine della compilazione una *web app* genererà l'etichetta. Anche i grafici saranno *data driven*, ossia generati dinamicamente con i dati forniti dal compilatore del form e facilmente aggiornabili. La parte di analisi delle immagini dovrà essere svolta precedentemente, o in un secondo momento, per poi essere inserita nel foglio illustrativo. La label di GEDE potrà poi essere inserita all'interno del proprio sito web tramite sistema di incorporazione (*embedding* o *link*).

L'utente viene a contatto con lo strumento grazie al sito web di presentazione progettato che introduce brevemente il tema, l'utilità di GEDE, le principali motivazioni e le istruzioni per creare l'etichetta. Il sito deve essere quanto più autoesplicativo, essendo l'unico strumento fornito per la comprensione dell'etichetta. Si compone di tre pagine principali: la landing page narra il problema e come GEDE si propone di risolverlo, una seconda pagina si focalizza sulle istruzioni per creare l'etichetta ed una terza pagina dove si specifica il sistema di analisi delle immagini, fornendo esempi per ogni tipologia di bias da individuare.

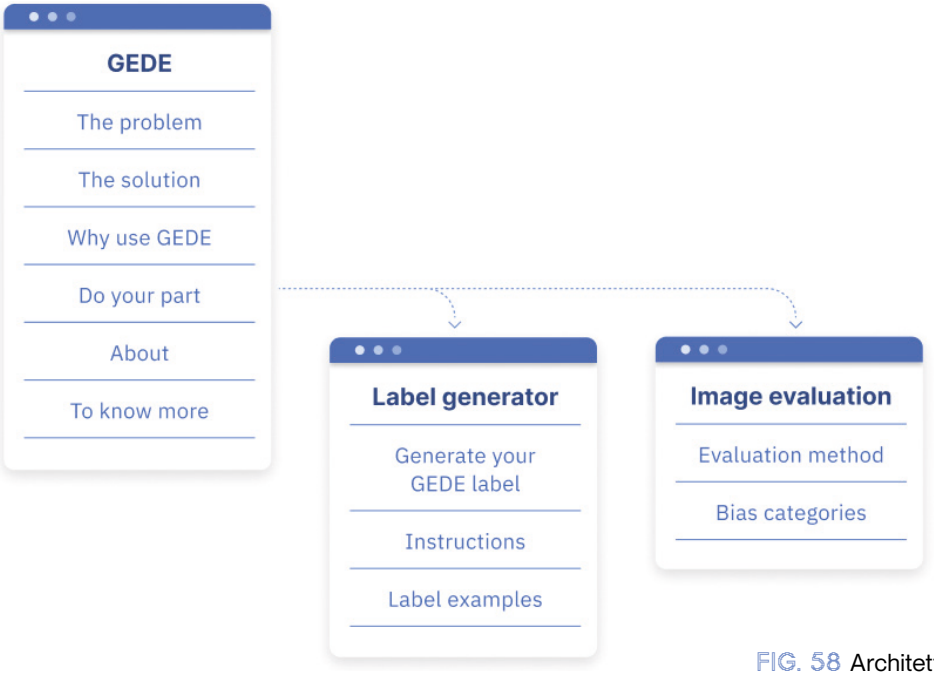


FIG. 58 Architettura delle informazioni del sito di GEDE.

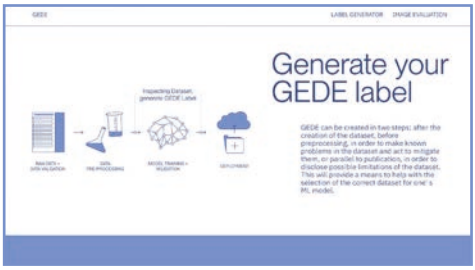


FIG. 59 Schermate del sito di presentazione di GEDE e Qr code.

5.4.2 Il design di GEDE

I parametri per il design visivo di GEDE sono stati definiti dall'identificazione del problema, dalla tipologia degli obiettivi ed dal criterio di valutazione delle alternative. Il problema del gender bias nell'IA è di ampio respiro, non si limita di certo ad un livello nazionale. Da qui deriva la scelta di progettare lo strumento in lingua inglese.

La possibilità di condividere l'etichetta e di incorporarla in qualsiasi sito è determinante per la scelta tipografica: il font deve essere gratuito ed open source. Scegliere un Google Font risulta la strada più sostenibile in quanto garantisce la compatibilità con tutti i dispositivi e una minima velocità di caricamento. La font utilizzata è IBM Plex™[10], famiglia di caratteri internazionali disegnata da Mike Abbink (IBM BX&D), in collaborazione con Bold Monday, una fonderia indipendente olandese. Plex è stato progettato specificatamente per catturare lo spirito e la storia di IBM e illustrare la fondamentale relazione tra uomo e macchina. Il risultato è un carattere Grotesque neutro e contemporaneo, ma non troppo serio, che include un Sans, Sans Condensed, Mono, Serif, e diversi altri stili per diverse lingue, e ha un'eccellente leggibilità nella stampa, nel web e nelle interfacce mobile.

9 Link: <ibm.com/plex/concept/>.

Sans	Mono	Serif
Thin	Thin	Thin
Thin Italic	*Thin Italic*	*Thin Italic*
Extralight	Extralight	Extralight
Extralight Italic	*Extralight Italic*	*Extralight Italic*
Light	Light	Light
Light Italic	*Light Italic*	*Light Italic*
Regular	Regular	Regular
Italic	*Italic*	*Italic*
Text	Text	Text
Text Italic	*Text Italic*	*Text Italic*
Medium	Medium	Medium
Medium Italic	*Medium Italic*	*Medium Italic*
Semibold	**Semibold**	**Semibold**
Semibold Italic	***Semibold Italic***	***Semibold Italic***
Bold	**Bold**	**Bold**
Bold Italic	***Bold Italic***	***Bold Italic***

FIG. 60 IBM Plex™ specimen.

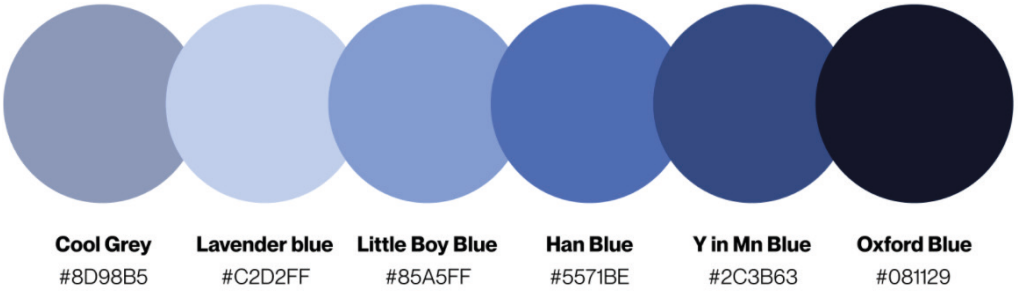


FIG. 61 Palette di GEDE: scala di toni di blu.

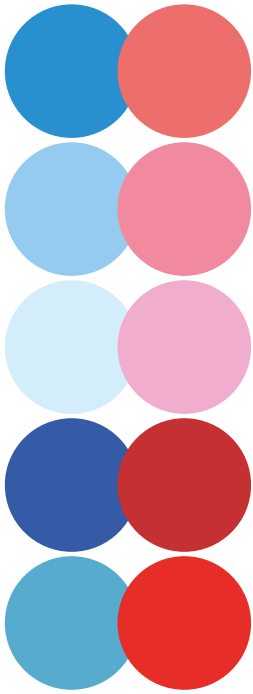
Si andrà ora a descrivere la semiosi, ovvero la correlazione tra la forma data al progetto e la sua percezione. Tale processo non ha riguardato solamente la fase di progettazione grafica nella quale tramite colori, forme, elementi grafici ecc, si è cercato di inserire i contenuti nell'ambiente più adatto alla loro lettura, ma i contenuti stessi sono stati redatti ponendo particolare attenzione al tono con cui questi avrebbero raccontato la vicenda. GEDE deve infatti essere credibile, rilevante e significativa, per questo l'approccio è assolutamente minimale ed evita qualsiasi tipo di enfasi.

Anche l'uso dei colori rispecchia in pieno l'approccio minimale della semiosi: tutto si colora di blu, ampiamente usato nel mondo tecnologico. Una scala di blu studiata secondo le regole della color blindness, e del contrasto, è utilizzata per rappresentare la scala dei 6 livelli di bias.

Molto *effort* è posto nella scelta del colore per rappresentare i diversi tag che identificano i sessi nel dataset. È noto infatti l'utilizzo da parte di moltissime visualizzazioni di dati del codice rosa e blu per femminile e maschile. Si vedano, tra gli esempi più recenti, visualizzazioni del New York Times, Bloomberg e il Wall Street Journal. Due progetti di The Pudding (2017) che analizzano il genere nei film, sottolineando le disuguaglianze, ricorrono al sistema rosso-blu e solo un loro terzo progetto (Gender Representation of Comic Book Characters) sperimenta leggermente, utilizzando il blu scuro per gli uomini e l'oro per le donne (cfr. Muth 2018).

C'è un forte dibattito sull'argomento: è chiaro che, utilizzando il sistema rosa-blu, il lettore sia in grado di decifrare il grafico in modo rapido, senza nemmeno necessitare

Rosa e blu nella visualizzazione dei dati:



The Pudding, 2017
She Giggles, He Gallops

Bloomberg, 2016
This chart shows who marries CEOs, Doctors, Chefs and Janitors

New York Times, 2017
The Words Men and Women Use When They Write About Love

DailyMail, 2018

ZEIT 2018

Fonte: Adattato da *Datawrapper, gender color 2018*

FIG. 62 La presenza del sistema rosa/blu nella visualizzazione dei dati.

di una legenda. Ma è anche vero che nella nostra cultura, come si è già potuto osservare nel corso della tesi, questi colori portano con sé tutto il bagaglio di stereotipi di genere. Il rosa indica il femminile, ma indica anche ragazze deboli e timide che giocano con le bambole; il blu indica il maschile, ma anche ragazzi forti e a volte violenti (cfr. Muth 2018). Creando un grafico con il rosa e il blu si avallano stereotipi di genere. Al contrario invece, grafici che capovolgono i colori stereotipati possono essere difficili da leggere: se i lettori infatti vedono rosa e blu in un grafico sulle distinzioni di genere, non saranno propensi a consultare la legenda e questo porterebbe a visualizzazioni potenzialmente fuorvianti.

Il *Telegraph* mette in campo un'alternativa: utilizza il viola per le donne e il verde per gli uomini. Fraser Lyness,

FIG. 63 Campagna "Votes For Women", inizio 20esimo secolo.

direttore del giornalismo grafico al Telegraph, spiega la scelta affermando che i colori sono ispirati dalla campagna "Votes for Women" nel Regno Unito, parte del movimento iniziale di suffragio all'inizio del 20esimo secolo. Apparentemente simboleggiano viola per la libertà e la dignità, bianco per la purezza, verde per la speranza.

> When deciding which gender aligned with which color, it was more a case of trying to prioritize women in the order of genders. Against white, purple registers with far greater contrast and so should attract more attention when putting alongside the green, not by much but just enough to tip the scales. In a lot of the visualisations men largely outnumber women, so it was a fairly simple method of bringing them back into focus. (Lyness)[10]

10 Fraser Lyness citato da Muth, 2018.

GEDE trae ispirazione dalle riflessioni del Telegraph, e riprende i due colori, rispettando i suggerimenti di Lyness. Si codifica, quindi: con il verde i tag riguardanti il maschile, con viola i tag riguardanti il femminile. Boy e Girl sono declinati con gli stessi colori in opacità e person è di colore magenta, colore primario.

La progettazione del sito web si fa carico di pensare a chi opererà dopo il designer, lo sviluppatore, per questo è necessario pensare ad un sistema che si compone tramite aspetti modulari, che facilitano la messa a terra del codice. L'inserimento di schede modulari ha il fine inoltre di movimentare la pagina e rendere la lettura più semplice, grazie ad un maggior coinvolgimento del lettore.

Il concetto alla base di GEDE origina dai ragionamenti fatti da D'Ignazio e Klein in Data Feminism. Questo anche per quanto concerne la visualizzazione dei dati. In particolare nell'etichetta ci si sofferma sul sesto principio per la visualizzazione femminista dei dati, ossia rendere visibile il lavoro (proprio e altrui) (cfr. D'Ignazio e Klein 2016). Il principio rende necessario dichiarare chi ha redatto l'etichetta, che risulterà quindi la prima domanda del *form*.

Il concetto si riferisce al "god trick", teorizzato da Donna Haraway: a livello epistemologico la nozione di *situated knowledge* consiste in uno sforzo per pensare al di fuori della dualità di oggettività-relativismo che è sia inefficace che dannosa per gli scopi femministi. Da un lato l'oggettività è stata compromessa in quanto solo apparentemente neutrale e di fatto sovraccarica di relazioni di potere. Inoltre, l'oggettività intesa come imparzialità e una «vista dall'alto, dal nulla» (god trick) è una prospettiva che sotto la maschera della neutralità, o del nulla (ma che abbraccia tutti), nasconde una posizione molto specifica (maschile, bianca, eterosessuale, umana) e quindi rende questa posizione fintamente universale.

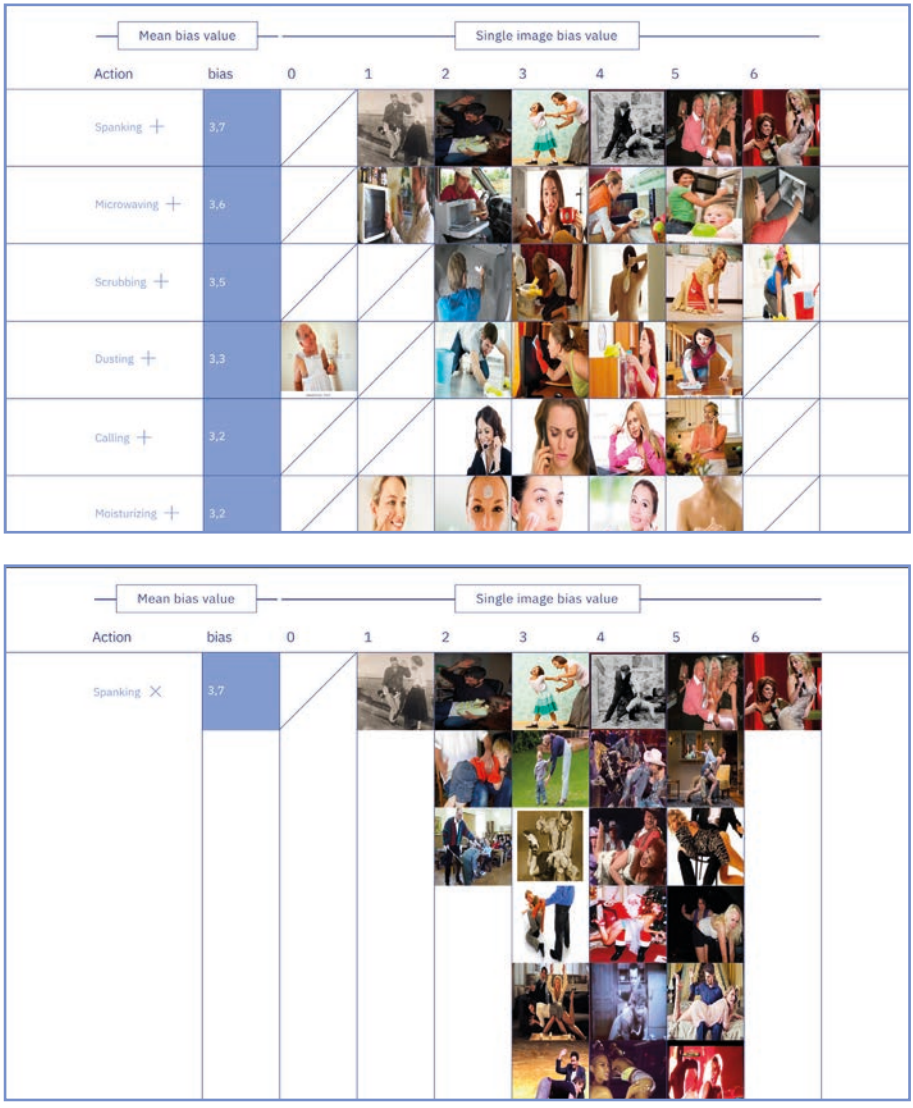
La visualizzazione delle immagini si ispira al metodo creato da Manovich in Media Visualization, basato su due step principali:

1. *feature extraction*: consiste nell'utilizzare i metodi delle immagini digitali per misurare un numero di caratteristiche visive (features). In questo processo, le caratteristiche visive sono mappate attraverso numeri, processo che in computer science è conosciuto come *feature extraction*.

2. vengono create visualizzazioni 2D che posizionano le immagini in base ai loro valori. Le immagini quindi, calcolate dal software nello step 1 diventano coordinate nello spazio 2D. In questo modo, le differenze tra le immagini sono tradotte nella loro dimensione nello spazio, diventando facilmente comprensibili all'occhio umano.

In aggiunta, il grafico è interattivo: all'ingresso è infatti fornita una sintesi che rappresenta i principali verbi/oggetti e le prime 6 immagini (una per valore) che compaiono. È possibile poi approfondire l'analisi e osservare la distribuzione di immagini per ogni verbo nei differenti valori e scoprire tutti i verbi/oggetti compresi nel campione di analisi.

FIG. 64 Catalogo di immagini nella terza sezione di GEDE.



GEDE, Gender Debiaser

imSitu

01 Overview | 02 Numbers | 03 Images

01 Overview

About

- Brief description:
  imSitu is a dataset supporting situation recognition, the problem of producing a concise summary of the situation an image depicts including: (1) the main activity, (2) the participating actors, objects, substances, and locations and most importantly (3) the role these participants play in the activity. The role set used by imSitu is derived from the linguistic resource FrameNet and the entities are derived from ImageNet. The data in imSitu can be used to create robust algorithms for situation recongntion.
- Purpose of creation:
  first dataset for situation recognition: generalized activity recognition and human-object interaction, using the assignment of roles to define how actors, objects, substances and locations participate in activities.

Who

- **Creator**: M. Yatskar, L. Zettlemoyer, A. Farhadi
- **Entity**: 1. Computer Science & Engineering, University of Washington, WA
  2. Allen Institute for Artificial Intelligence, Seattle, WA
- **Funder**: N.D.
- **Associated grant** (if available): N.D.
- **Model Developer**: N.D.

Data collection

- Type of data acquisition:
  verbs from FrameNet, images gathered from Google image search with query expansion techniques and labeled with complete situations on Amazon Mechanical Turk.
- Was data validated/verified? How?
  Workers were instructed to select images that (1) are not modified or computer generated and (2) contain at least some part of the main entity doing the action
  in the image. All images were annotated by three crowd workers. Both *Situation Recognition* and *Activity and Object Recognition* were evaluated with quality control algorithms.
- Entities involved in data collection:
  team; Mechanical Turk annotators; five computer science undergraduates.
- Type of compensation (annotators):
  annotation cost approximately $80 per verb.

Badges

QUALITY REVIEW | AGGREGATED HUMAN DATA | ABOUT HUMANS
NO ETHICAL REVIEW | N.S. UPDATE FREQUENCY | NON FOR PROFIT FUNDING SOURCE
MULTI SOURCE DATA | SENISIBLE DATA | INFO ABOUT SUBPOPULATIONS

Alerts

N.D.

Use Cases

- Situation Recognition.
- Study of objects, of activities, and their interactions through semantic roles.

Privacy

Open data

Type of content

JSON, Py

Dates

- **Dataset creation date**: 2016
- **Last dataset update date**: 2019
- **Model creation date**: N.D.

Sources

- Dataset origin link.
- Dataset paper link.

02 Numbers

Stacked Chart

Total number of pictures with tags*: #Person #Man #Woman #Boy #Girl

Hover the dataviz to discover more

*Dataset tags. Note: no data for nonbinary people.

Donut Chart

Total number of pictures per verbs with tags: #Man #Woman

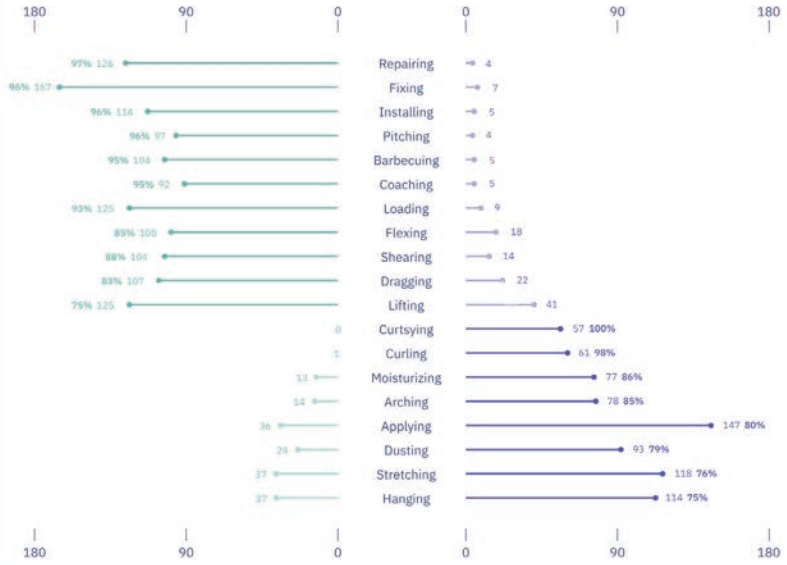
Donut Chart

Total number of pictures per verbs with tags: #Man #Woman

Butterfly Chart

Number of pictures per verb with tag: #Man #Woman

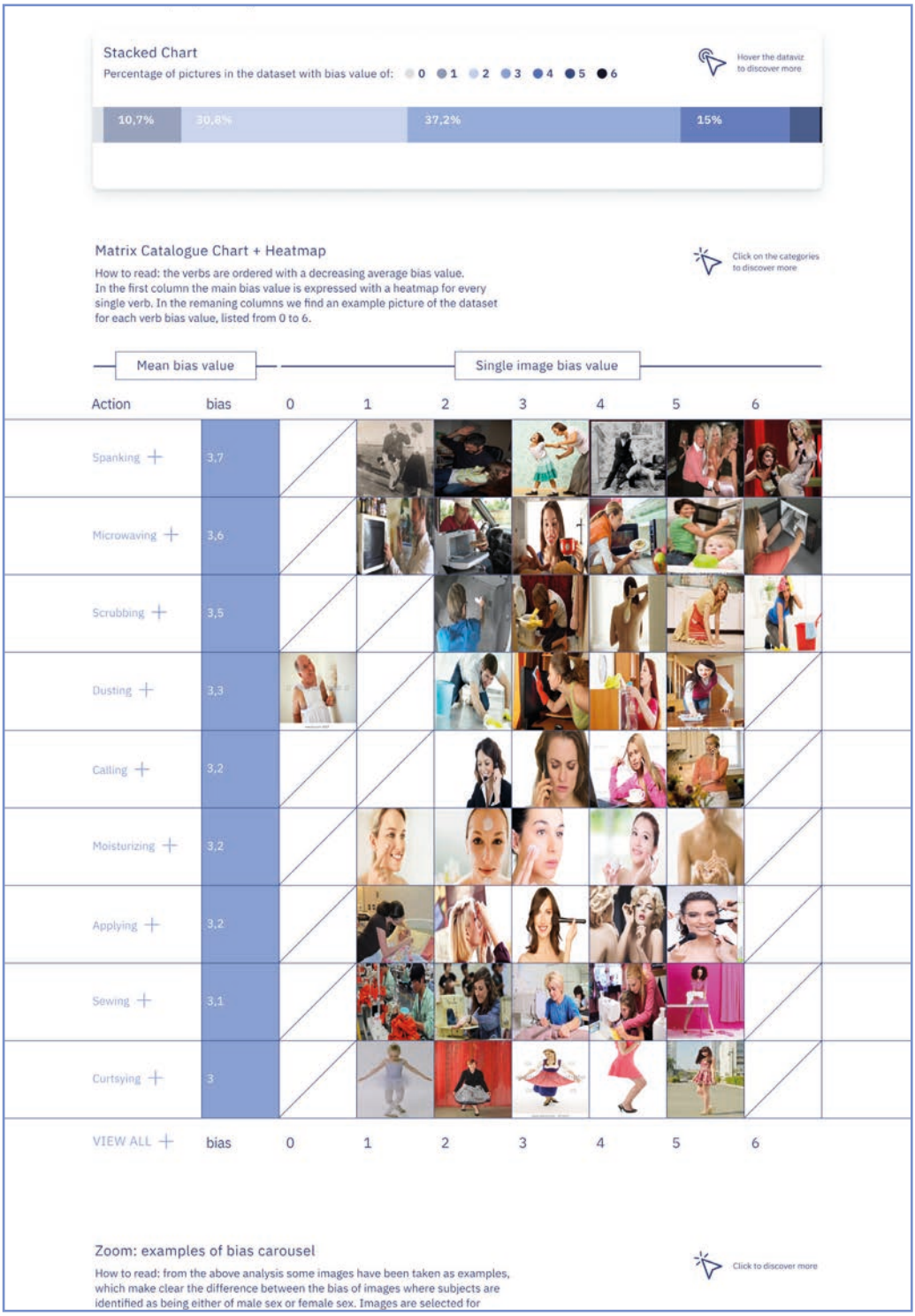
The percentage of the highest population is reported beside the number of images.



03 Images

3/6 bias index

Meaning: the average photo has at least 3 bias points present.
The bias point is attributed if at least one stereotypical element is found in the photo, according to the evaluation form.


Stacked Chart
Percentage of pictures in the dataset with bias value of: 0 1 2 3 4 5 6
Hover the dataviz to discover more
10,7%
30,8%
37,2%
15%
Matrix Catalogue Chart + Heatmap
How to read: the verbs are ordered with a decreasing average bias value.
In the first column the main bias value is expressed with a heatmap for every single verb. In the remaning columns we find an example picture of the dataset for each verb bias value, listed from 0 to 6.
Click on the categories to discover more
Mean bias value
Single image bias value
Action
bias
0
1
2
3
4
5
6
Spanking
3,7
Microwaving
3,6
Scrubbing
3,5
Dusting
3,3
Calling
3,2
Moisturizing
3,2
Applying
3,2
Sewing
3,1
Curtsying
3
VIEW ALL
bias
0
1
2
3
4
5
6
Zoom: examples of bias carousel
How to read: from the above analysis some images have been taken as examples, which make clear the difference between the bias of images where subjects are identified as being either of male sex or female sex. Images are selected for
Click to discover more

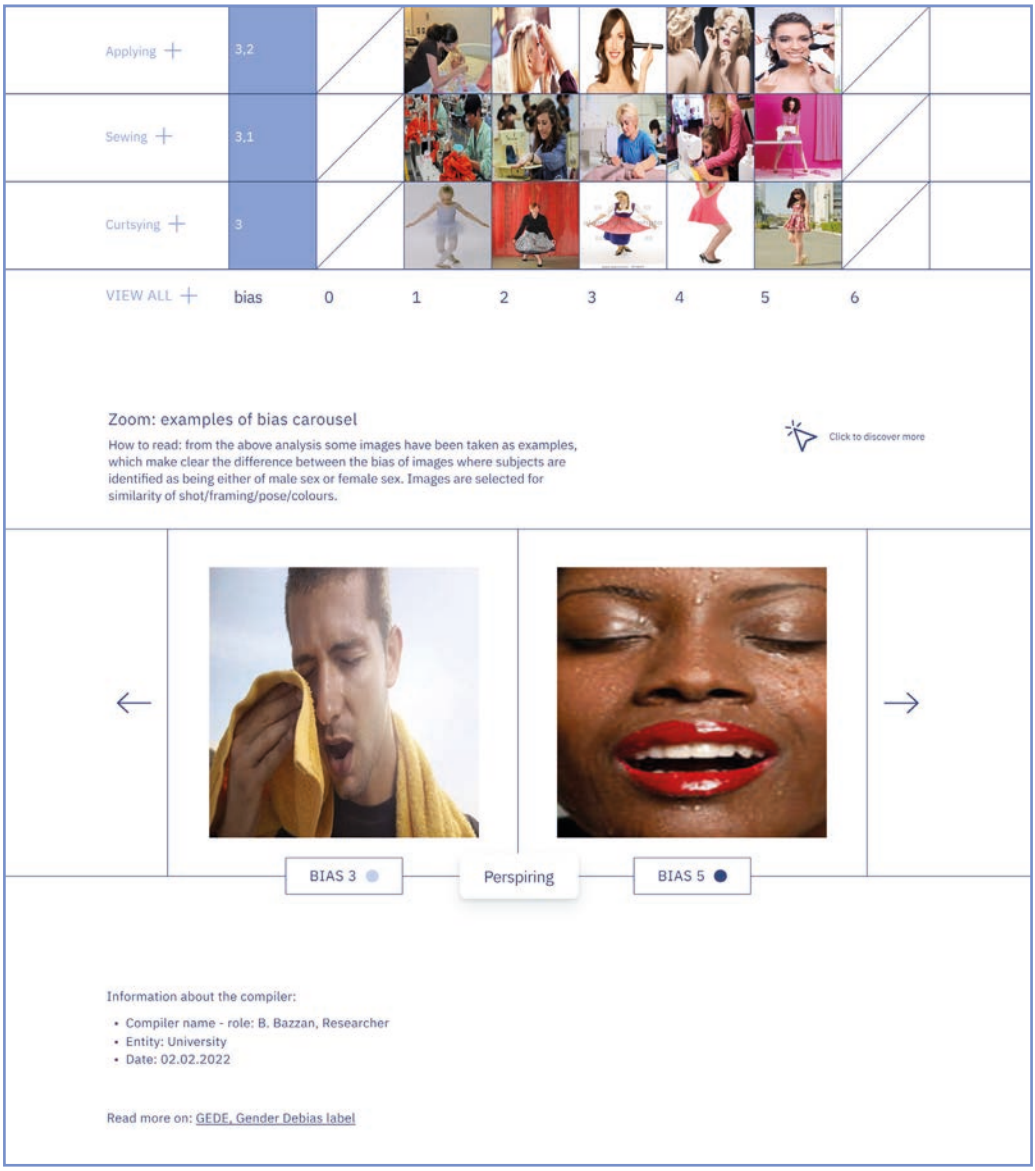


FIG. 65 GEDE per imSitu.

GEDE, Gender Debiaser

Open Images V6

01 Overview | 02 Numbers | 03 Images

01 Overview

About

- Brief description:
  Open Images is a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives.
  The dataset is split into a training set (9,011,219 images), a validation set (41,620 images), and a test set (125,436 images). The images are annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives as described below.
- Purpose of creation:
  image classification, object detection and visual relationship detection.

Who

- **Creator:** Krasin I., Duerig T., Alldrin N., Ferrari V., Abu-El-Haija S., Kuznetsova A., Rom H., Uijlings J., Popov S., Kamali S., Malloci M., Pont-Tuset J., Veit A., Belongie S., Gomes V., Gupta A., Sun C., Chechik G., Cai D., Feng Z., Narayanan D., Murphy K.
- **Entity:** Google Cloud Vision API
- **Funder:** Google

Data collection

- Type of data acquisition:
  The images have been collected from Flickr without a predefined list of class names or tags, leading to natural class statistics and avoiding an initial design bias. The set of classes included in the Open Images Dataset is derived from JFT, an internal dataset at Google with millions of images and thousands of classes (Hinton et al., 2014; Chollet, 2017; Sun et al., 2017). W
- Was data validated/verified? How?
  - Extract relevant metadata of all images to give proper attribution
  - Remove images containing inappropriate content (porn, medical, violence, memes, etc.) using the safety filters on Flickr and Google SafeSearch.
  - for the annotation a computer-assisted protocol is used. First they apply an image classifier to generate candidate labels for all images and then ask humans to verify them.
- Entities involved in data collection:
  Google Cloud Vision API, human annotator froom a Google-internal pool and from a crowdsource external pool
- Type of compensation (annotators): N.D.

Badges

QUALITY REVIEW | AGGREGATED HUMAN DATA | ABOUT HUMANS
ETHICAL REVIEW | YEAR ANNUAL UPDATE FREQUENCY | FOR PROFIT FUNDING SOURCE
SINGLE SOURCE DATA | SENISIBLE DATA | NO INFO ABOUT SUBPOPULATIONS

Alerts

N.D.

Use Cases

- Large-scale multi-label and multi-class image classification

Privacy

Open data, Public dataset

Type of content

JSON, Py

Dates

- **Dataset creation date:** 2017
- **Last dataset update date:** 2021

Sources

- Dataset origin link.
- Dataset paper link.

02 Numbers

Stacked Chart

Total number of pictures with tags*: #Person #Man #Woman #Boy #Girl

Hover the dataviz to discover more

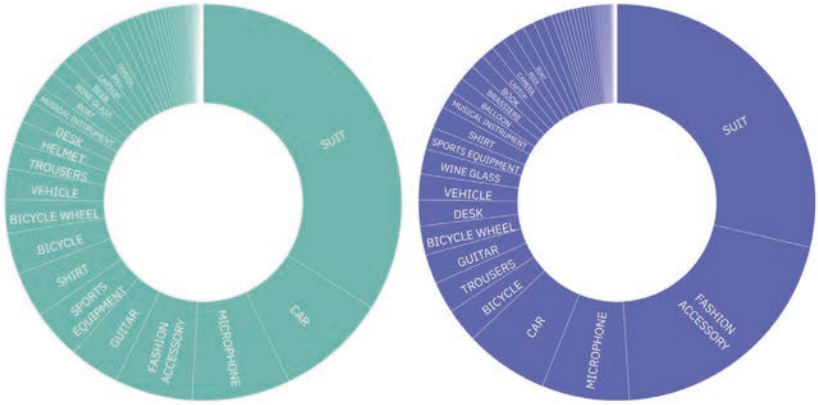
*Dataset tags. Note: no data for nonbinary people.

Donut Chart

Total number of pictures per verbs with tags: #Man #Woman

Donut Chart

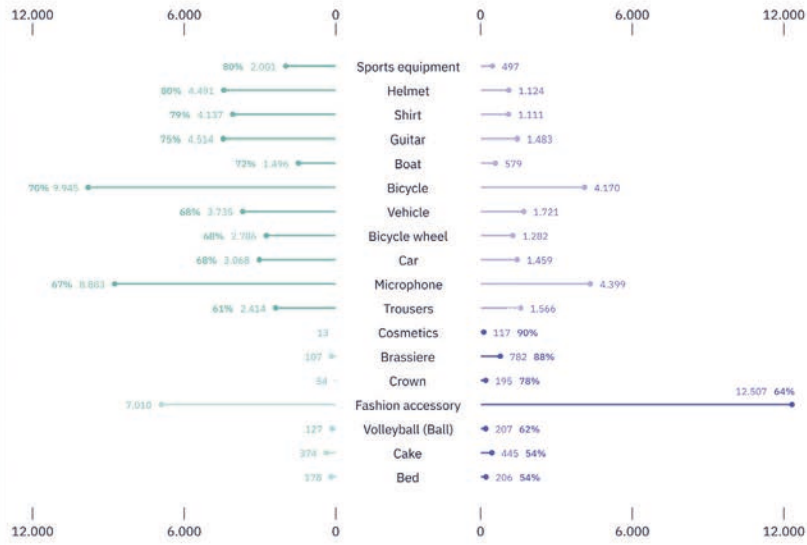
Total number of pictures per verbs with tags: #Man #Woman



Butterfly Chart

Number of pictures per verb with tag: #Man #Woman

The percentage of the highest population is reported beside the number of images.



03 Images

2,5/6 bias index

Meaning: the average photo has at least 2,5 bias points present. The bias point is attributed if at least one stereotypical element is found in the photo, according to the evaluation form.


Stacked Chart
Percentage of pictures in the dataset with bias value of: 0 1 2 3 4 5 6
Hover the dataviz to discover more
6,9
20
28
23,1
14,66
5,5
Matrix Catalogue Chart + Heatmap
How to read: the verbs are ordered with a decreasing average bias value. In the first column the main bias value is expressed with a heatmap for every single verb. In the remaning columns we find an example picture of the dataset for each verb bias value, listed from 0 to 6.
Click on the categories to discover more
Mean bias value
Single image bias value
Object
bias
0 1 2 3 4 5 6
Washing Machine
4,4
Billboard
4,0
Brassiere
3,8
Mobile Phone
3,5
Cosmetics
3,4
Drill (Tool)
3,4
Sword
3,3
Crown
3,1
Fashion Accessory
3,1
VIEW ALL
bias
0 1 2 3 4 5 6
Zoom: examples of bias carousel
How to read: from the above analysis some images have been taken as examples, which make clear the difference between the bias of images where subjects are identified as being either of male sex or female sex. Images are selected for similarity of shot/framing/pose/colours.
Click to discover more

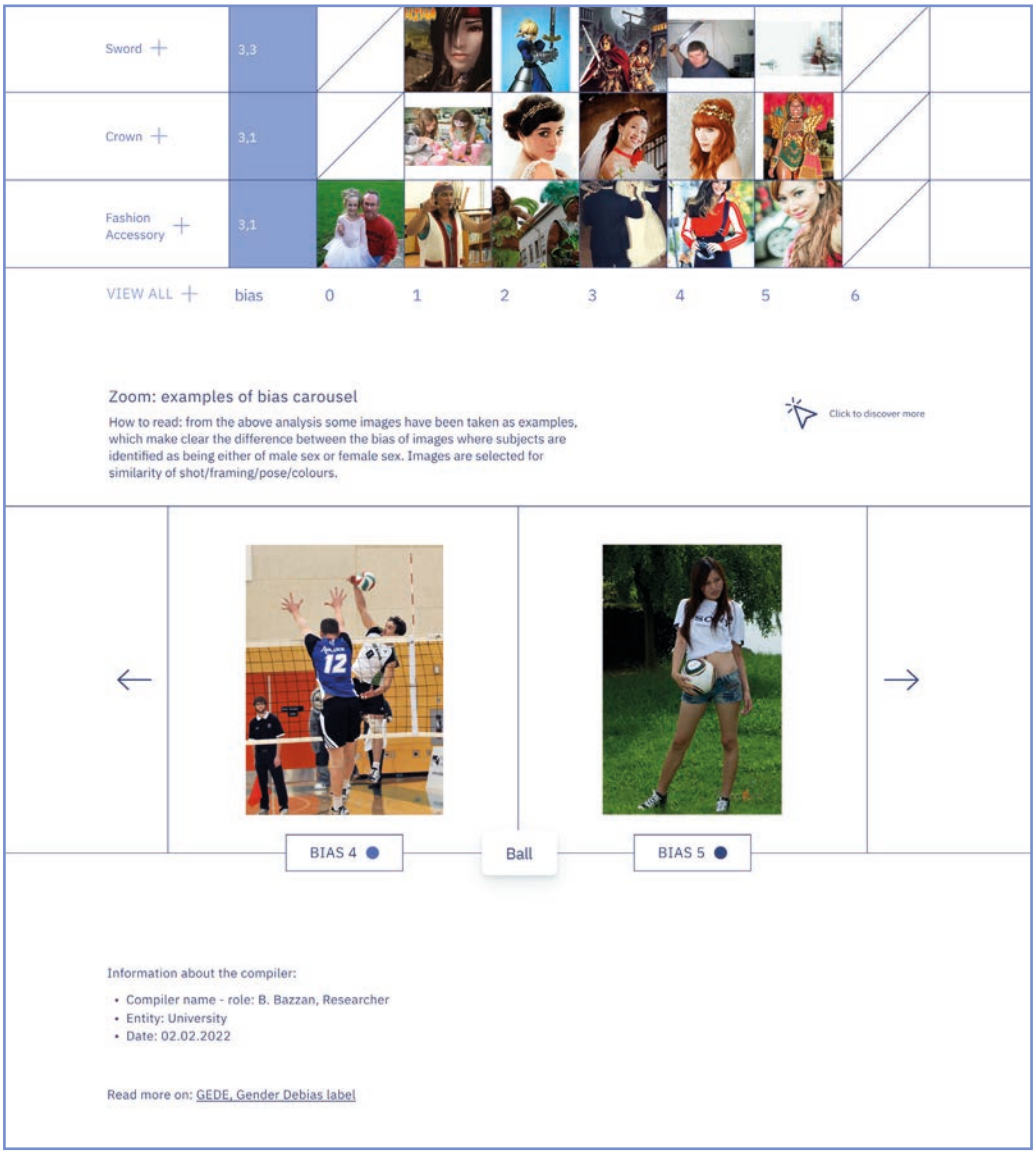


FIG. 66 GEDE per Open Images v6.

5.5 Gli insights

GEDE fornisce una *overview* del dataset, dalla quale è possibile trarre conclusioni e insight. Si nota, ad esempio, che i soggetti con caratteristiche fisiche maschili vengono di sovente ritratti con braccia incrociate ed espressioni serie o severe e, al contrario, i soggetti con caratteristiche fisiche femminili generalmente appaiono sorridenti e amichevoli. Il pattern ricalca lo stereotipo che:

> Come ipotizzano Hugenberg e Sczesny (2006)[...] vuole che le donne sorridano di più è lo stesso che facilita la percezione di emozioni felici coerenti con lo stereotipo nei visi femminili. "Lo studio rivela che, in un compito di categorizzazione dei volti mostranti felicità o rabbia, i volti felici vengono categorizzati più velocemente ed accuratamente su volti target femminili piuttosto che maschili e l'inverso per la rabbia (Hugenberg e Sczesny, 2006)." La congruenza tra emozioni espresse dal volto e stereotipi di genere sui ruoli sociali incide quindi sulla riconoscibilità del genere, ma anche sulla percezione sociale dell'emittente. (De Piccoli e Rollero, 2018)

Le rare volte in cui non le donne sono ritratte sorridendo sono etichettate come "assillanti", che si lamentano continuamente e necessitano dell'attenzione dei maschi.

FIG. 67 Immagini estratte dai verbi *reading*, *telephoning*. (imSitu)

Affine a questo tropo è quello della violenza linguistica, in quanto le donne sono non di rado classificate come «bitchy» (cfr. Lukas 2002). Lo stereotipo è ricalcato dal dataset IS: il verbo *nagging* (inglese per assillare) contiene 57 fotografie con tag *woman*, 18 *wife*, 9 *mother* e solo 8 *man*.

Gli stereotipi nelle occupazioni sono ripetute alla perfezione da entrambi i dataset, dove infermiera e casalinga rimangono occupazioni fortemente associate al mondo femminile, mentre dottore, manager e atleta all'opposto mondo maschile. *Teaching* compare in IS 37 volte con il tag *woman* e 17 con il tag *man*. Sebbene compaiano anche uomini nelle immagini relative all'insegnamento, in entrambi i dataset insegnanti uomini e donne sono rappresentati in modi molto differente. I primi infatti si legano ad ambienti più adulti, quali scuole superiori o università, mentre le seconde sono prevalentemente collocate in scuole elementari o medie.

FIG. 68 Immagini tratte dal verbo *educating* (imSitu) e dall'oggetto *whiteboard* (OpenImages v6).

Anche gli stereotipi negli oggetti e negli sport sono ampiamente richiamati, vediamo infatti un'affollamento di figure maschili in *coaching* (92) e *flexing* (100) e femminili in *volleyball (ball)* (207 a 127), sempre considerato uno sport femminile. Martin 2018, in "What Happens When We Give Everything a Gender" afferma che «when one sees a human-connected entity as gendered (such as believing fo-

otball is masculine), it likely evokes many stereotypes about men as rough and strong, reifying and strengthening masculine stereotypes about men, making these qualities seem natural» e continua affermando che «Our research found that gendering human-connected entities—like toys—does in fact increase stereotyping and bias that reinforce gender inequality» (Martin 2018).

La genderizzazione degli oggetti è un ennesimo tropo fortemente trattato dalle pubblicità, dove le femmine sono più spesso mostrate nelle pubblicità di prodotti per la pulizia, prodotti alimentari, prodotti di bellezza, farmaci, abbigliamento ed elettrodomestici. I maschi sono più spesso mostrati in annunci per automobili, viaggi, bevande alcoliche, sigarette, banche, prodotti industriali, mezzi di intrattenimento e aziende industriali (cfr. Courtney e Lockeretz 1979: 92-95).

> La pubblicità partecipa alla socializzazione di genere differenziando i messaggi che i piccoli/e telespettatori/trici ricevono. Così, ai maschi vengono proposti principalmente giochi di avventura, armi, ambientazioni in campo aperto che rimandano a mondi fantastici attraverso parole con cui li invitano a "conquistare", "entrare in azione", "combattere", "difendere". Alle femmine, invece, si propone di accudire, pulire, cucinare, truccarsi. Le ambientazioni pubblicitarie sono sempre in un interno, le voci di sottofondo sono morbide e rassicuranti. Al grido di "è facile", o "è fashion!" scorrono immagini di forni, carrelli delle pulizie, bambolotti, trucchi, bambole alla moda. (Dulbecco 2020)

Ecco che in OI a "cosmetics" corrisponde 117 tag femminili e 13 maschili. ed in IS per "dusting" 24 sono i tag maschili e 93 i femminili, evidenziando la forte connessione tra femminile e superficialità o ancora femminile e lavori domestici, verità che sembra indiscutibile in entrambi i corpi di dati analizzati.

OI ha un bias value un po' più basso (2,5 rispetto ad un 3 di IS), forse dovuto all'aggiornamento più costante dei dati e avvenuto in tempi più recenti rispetto ad IS. Può quindi avere un'attenzione maggiore verso temi che stanno emergendo ora. Anche OI presenta tuttavia significativi stereotipi, di cui di seguito ne verrà analizzato qualcuno.

Nel dataset le figure maschili sono spesso rappresentate nello svolgimento dell'azione, mentre le figure femminili sono ritratte prima o dopo esempio aver eseguito l'azione, ad esempio aver suonato o ballato. La scelta codifica un significato: i personaggi non impegnati sembrano non attivi, sono passivi davanti all'obiettivo fotografico.

FIG. 69 (Pagina destra) Immagini tratte dagli oggetti *ski, ball, racket, sports equipment, musical instrument.* (OpenImages v6).

Per quanto riguarda le categorie di oggetti, è stato notato che esiste *Brassiere*, parola tratta dal francese che indica il reggiseno femminile, ma non esiste alcuna classe per descrivere intimo maschile, quale "male underwear", un generico "underwear" o "lingerie". Questo porta chiaramente ad avere molte più situazioni in cui le figure femminili sono poco o non vestite, aumentando i casi in cui si verifica la categoria di body display.

Un'altra osservazione riguarda l'uso, in particolare di OI della categoria *Girl* anche per indicare una donna in età adulta. Il fenomeno può essere analizzato osservando i dati nella sua interezza, anche solo valutando che al tag *woman* corrispondono circa 61.400 immagini e al tag *girl* circa 16.800, in comparazione con al 110.517 *man* e 6.500 *boy*.

FIG. 70 (Pagina destra) Immagini tratte dagli oggetti *washing machine, mobile phone.* (OpenImages v6).

Ci sono poi ancora richiami a stereotipi nelle occupazioni: ad esempio in OI, analizzando le statistiche vediamo che per il verbo *handshake* ci sono *man* 33: e 1 *woman*. Anche qui si sottolinea la connessione tra il femminile e la superficialità e l'estetica: sono 7010 le immagini di "fashion accessory" con tag *man*, a confronto delle 12.507 *woman*.

Sovente si vedono modelle di origini presumibilmente orientali a presentare prodotti, soprattutto tecnologici, quali lavastoviglie, smartphone o computer. Questo pattern sottolinea sia la tendenza all'esotizzazione sia la diversa composizione e origine geografica delle immagini.


SERIES
4G LTE

5.6 I limiti (delle piattaforme e dei dataset stessi)

> Usare il sesso biologico di qualcuno per caratterizzare in modo essenziale la personalità di un intero gruppo di persone è come operare chirurgicamente con un'ascia. Non è abbastanza preciso per fare del bene, probabilmente causerà molti danni. (Schmitt 2017; traduzione mia[11])

11 Citazione originale: «Using someone's biological sex to essentialize an entire group of people's personality is like surgically operating with an axe. Not precise enough to do much good, probably will cause a lot of harm» (Schmitt 2017).

Progettando GEDE sono emersi dei vincoli e problemi non facilmente risolvibili o arginabili.

Uno di questi riguarda la classificazione di genere: nei dataset si limita sempre ad un sistema binario, mai estendendosi in uno spettro. Considera solo la categoria uomo e donna o maschio e femmina (man-woman / male-female).

> Il sistema del binarismo e le combinazioni che ne conseguono sono però una griglia interpretativa che descrive la realtà in maniera incompleta. [...] All'interno della comunità LGBTQIA+ si dice spesso che "il genere è uno spettro", come lo spettro delle onde elettromagnetiche: significa che non esistono solo un genere femminile e un genere maschile, ma uno spettro continuo di generi tra questi due estremi. (Il Post 2021: 13-15)

I ricercatori Scheuerman, Paul e Brubaker (2019) riportano che, nei dataset analizzati, mentre etichette gender-neutral erano previste per certe immagini (quali "person" e "human") il genere era sempre presente nei servizi di riconoscimento del volto che usano la classificazione di genere (tutti tranne Google), perciò quando sono richieste etichette di genere esplicite (cfr. Scheuerman, Paul e Brubaker 2019). E, nel classificare il genere, i progettisti dei sistemi che hanno analizzato hanno scelto di utilizzare solo le categorie binarie. Le classificazioni sono poi registrate, misurate, etichettate e archiviate e poi spesso ne viene fatto uso commerciale, sotto forma di servizi basati su cloud, fornendo un'infrastruttura che terze parti possono utilizzare per creare o aumentare i propri servizi. Nel processo, questi servizi

propagano una visione riduzionista del genere fornita dall'infrastruttura sottostante. L'auto-identità non è usata dai sistemi di visione del computer. Dopo tutto, non può essere vista (cfr. Scheuerman, Paul e Brubaker 2019).

La classificazione binaria costringe gli utenti non binari a conformarsi alle aspettative cisnormative sul genere e ad adattarsi al bivio di genere demografico (M o F). è chiaro che i generi non-binari rappresentano una sfida per la classificazione di genere, ma allo stesso modo evidenziano le problematiche della progettazione di sistemi che trattano dati sensibili.

Il progetto di tesi tenta di passare ad un sistema più inclusivo, dove il classificatore di genere è specificatamente l'apparenza di genere: figure in apparenza femminili e figure in apparenza maschili. Rappresenta un cambiamento potenziale nella classificazione di genere, che si muove da una categorizzazione fissa biologicamente associata da una categoria biologicamente associata ad una qualità percepita. Lo stesso metodo è utilizzato in questa caso, dove non è previsto di dimostrare il genere dell'individuo, ne tantomeno si classifica il sesso biologico o l'identità di genere, quanto piuttosto l'aspetto esteriore del genere, che non fa parte di un aspetto culturale ma biometrico. Sebbene anche supporre l'aspetto di una persona potrebbe essere considerato sbagliato, specialmente in caso di persone non binarie, androgine, transessuali o che comunque non si identificano con il sistema binario.

> L'identificazione di genere, in particolare imposta dallo stato, è già stata usata per sorvegliare le identità trans (per esempio, impedendo agli individui trans di accedere all'assistenza sanitaria). Danni sociali e fisici potrebbero essere perpetrati usando la tecnologia della visione artificiale agli individui trans, che già affrontano alti livelli di molestie e violenza. La classificazione binaria del genere nell'analisi facciale potrebbe essere usata per ostacolare intenzionalmente l'accesso agli spazi sociali (ad esempio i bagni), limitare il movimento (ad esempio la Transportation Security Administration (TSA) degli Stati Uniti), e persino attuare una violenza sistematica e mirata se adottata da governi virulentemente anti-trans. (Scheuerman, Paul, e Brubaker 2019; traduzione mia)

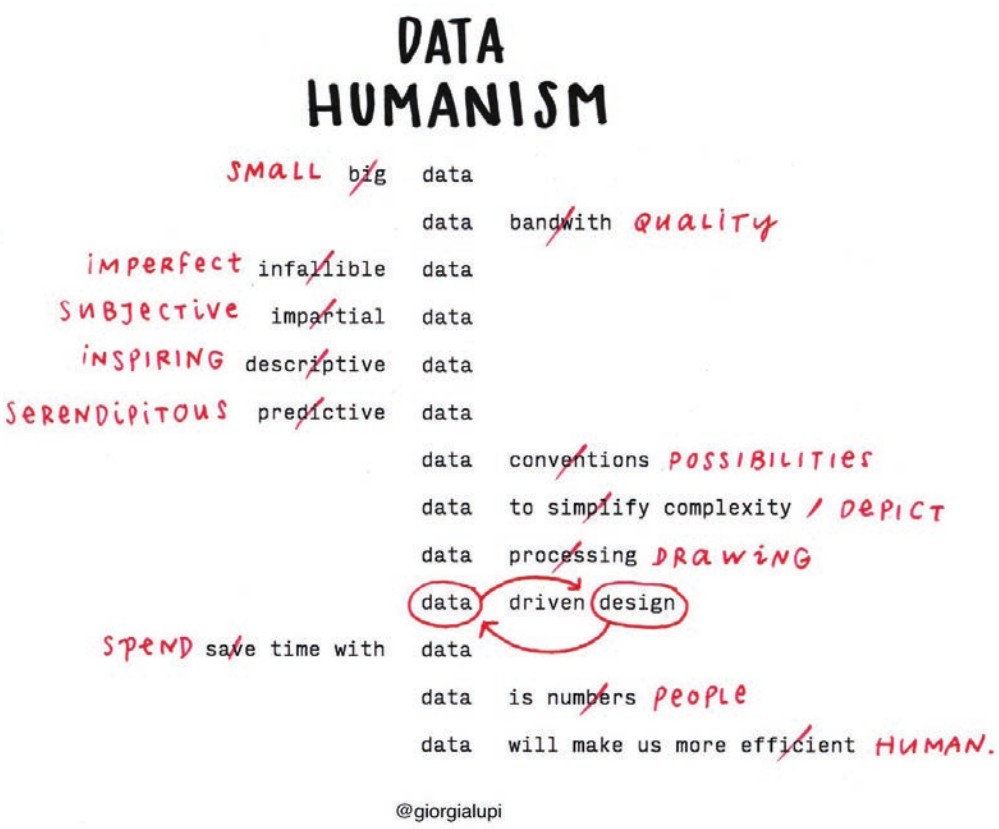


FIG. 71 Giorgia Lupi *Data Humanism, The Revolution will be Visualized* (2017).

Un ulteriore limite consiste nell'impossibilità di compiere un'analisi sul totale delle immagini: gli algoritmi, per funzionare, necessitano di una quantità ingente di dati, non analizzabili dall'occhio umano. Per questo risulta necessario, per compiere l'analisi qualitativa delle immagini, scegliere un campione più o meno grande che consenta al ricercatore di analizzare le immagini secondo i metri di valutazione prescritti. Si può però ribattere considerando il concetto di data humanism introdotto da Giorgia Lupi, information designer. Considerare i dati nel loro contesto, nella loro natura e organizzazione, stabilendo un modo per analizzarli. Per questo è risultato necessario analizzare la parte umana del dataset, quindi i comprendere come i tag di genere sono trattati, quali sono i numeri di genere e come vengono rappresentati dalle immagini incluse nel dataset. Goffman nella sua stessa analisi osserva che:

> Le complicazioni risiedono nel fatto che posare per una pubblicità porta sempre con sé uno strascico di sessualità, dato che le modelle appaiono come figure femminili e i modelli come figure maschili. (Goffman 2015: 125)

Quindi, egli sostiene che «nelle affermazioni sugli stereotipi sessuali, allora, troviamo una speciale garanzia per la ricaduta su riferimenti semplificati» (Goffman 2015: 125).

6.0 Conclusioni

AI system could help to discover existing inequality that might have remained hidden otherwise.
Borgesius 2018

6.1 Future implementazioni possibili: domanda aperta

L'obiettivo della tesi è sviluppare e applicare conoscenze di analisi visiva ai sistemi algoritmici, consentendo trasparenza nel contenuto dei dataset di allenamento. La ricerca ha dimostrato che è fondamentale riuscire a generare principi base che mirano all'inclusione di genere nelle infrastrutture della computer vision, poiché moltissimi sono gli esempi di ingiustizie derivate da dataset distorti, dalla assenza di dati di genere o dalla sotto o sovra-rappresentazione del genere in alcune arene sociali.

Nella strutturazione di GEDE è emerso che la commistione tra analisi di dataset e studi di genere non è ancora stata approfondita. È stato per questo necessario fare riferimenti a letterature ibride e multidisciplinari, che approfondissero da un lato il problema in termini di data science e dall'altro lato in termini di analisi visiva delle immagini. Il designer della comunicazione non è esente dal problema del gender bias nelle IA: è coinvolto in ogni fase della progettazione, dalla scelta delle immagini alla scelta delle variabili che il dataset propone, alla creazione dell'interfaccia del successivo modello algoritmico.

GEDE si propone come un modulo aggiuntivo alle etichette che individuano in termini statistici e matematici il contenuto del dataset, e sarebbe per questo ottimale una collaborazione diretta tra designer e specialisti del campo per progettare una formula in grado di unire i due moduli e fornire un'analisi complessiva del dataset di riferimento.

La ricerca e il tool sviluppato vanno considerati come un mero punto di partenza per ulteriori sperimentazioni e progressi. Un successivo step di avanzamento per il progetto potrebbe essere la sua sperimentazione. La procedura consisterebbe innanzitutto nel selezionare un algoritmo sviluppato e il suo dataset di allenamento, per poi applicarvi GEDE ed ottenere così il suo livello di gender bias.

A quel punto bisognerebbe correggere il dataset (sostituendolo con un nuovo dataset più "pulito", oppure aggiustando il campione di immagini) con l'obiettivo di partire da un dataset con un minor livello di gender bias. Infine l'algoritmo dovrebbe essere ri-allenato sul nuovo dataset. A questo punto un'analisi sui nuovi risultati forniti dall'algoritmo dovrebbe stabilire se effettivamente si è ottenuto un miglioramento, ottenendo i risultati più equi e con meno discriminazioni rispetto al suo funzionamento precedente. Questo dimostrerebbe sia l'efficacia del metodo di etichettatura, sia il principio stesso su cui si basa, ovvero che se si vuole giungere ad algoritmi più equi bisogna partire dai dataset di allenamento.

Un altro step potenzialmente da approfondire riguarda l'etichettatura manuale delle immagini, che risulta complessa e dispendiosa in termini di tempo. Si potrebbe automatizzare la valutazione delle immagini tramite la creazione di un algoritmo che identifica le varie tipologie di bias.

Si è inoltre rivelato fondamentale, nel corso della ricerca, la necessità di avere a supporto leggi e politiche che determinino gli standard di inclusione per l'uso del genere nella computer vision. Eventuali regolamentazioni potrebbero fare leva su tool come GEDE, che diventerebbero fondamentali per scegliere in modo più consapevole i dataset di allenamento, con il fine di ottenere algoritmi più giusti.

Fonti

Abbatecola, Emanuela, e Luisa Stagi, a c. di
2017 *Pink is the new black: Stereotipi di genere nella scuola dell'infanzia.* Rosenberg & Sellier. <doi.org/10.4000/books.res.4876>.

Abele, Andrea E., e Bogdan Wojciszke
2014 «Communal and Agentic Content in Social Cognition». In *Advances in Experimental Social Psychology*, 50:195–255. Elsevier. <doi.org/10.1016/B978-0-12-800284-1.00004-7>.

ABOUT ML, a c. di
2020 «Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles». *ABOUT ML*. Consultato il 26 luglio 2021, <partnershiponai.org/about-ml-get-involved/#read>.

Aiello, Giorgia
2016 «Taking Stock». *Ethnography Matters, The Person in the (Big) Data*. Consultato il 30 agosto 2021, <ethnographymatters.net/blog/2016/04/28/taking-stock/>.

2019 «Taking Stock: Researching Generic Images across Representation, Circulation and Recontextualization - 1 May 2019». Evento presentato al *How can we study the meanings and circulation of generic images online?*, University of Leeds. Consultato il 29 ottobre 2021, <kcl.ac.uk/events/taking-stock-researching-generic-images-across-representation-circulation-and-recontextualization>.

2020 «Visual Semiotics: Key Concepts and New Directions». In *The SAGE Handbook of Visual Research Methods*, di Luc Pauwels e Dawn Mannay, 367–80. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications, Inc. <doi.org/10.4135/9781526417015.n23>.

Ananny, Mike, e Kate Crawford
2018 «Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability». *New Media & Society 20* (3): 973–89. Consultato il 25 giugno 2021, <doi.org/10.1177/1461444816676645>.

Anderson, Chris
2008 «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete». *Wired, Science*, giugno. Consultato il 26 giugno 2021, <wired.com/2008/06/pb-theory/>.

Arnold, Matthew, Rachel K. E. Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilovic, Ravi Nair, et al.
2019 «FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity». Consultato il 11 giugno 2021, <arxiv.org/abs/1808.07261>.

Balakrishnan, Tara, Michael Chui, Bryce Hall, e Nicolaus Henke
2020 «Global Survey: The State of AI in 2020». Survey di *McKinsey Digital*. Consultato il 26 giugno 2021, <mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020>.

Barbieri, Davide, Jakub Caisl, Marre Karu, Giulia Lanfredi, Blandine Mollard, Vytautas Peciukonis, Maria Belen Pilares La Hoz, Jolanta Reingarde, e Lina Salanauskaite
2020 *Gender Equality Index 2020: Digitalisation and the Future of Work*. A cura di European Institute for Gender Equality. <op.europa.eu/publication/manifestation_identifier/PUB_MHAF20001ENN>.

Barocas, Solon, e Andrew D. Selbst
2016 «Big Data's Disparate Impact». *SSRN Electronic Journal*. Consultato il 26 giugno 2021, <doi.org/10.2139/ssrn.2477899>.

Barsan, Ilinca
2020 «Research Reveals Inherent AI Gender Bias». *Wunderman Thompson, Insight*. Consultato il 1 ottobre 2021, <wundermanthompson.com/insight/ai-and-gender-bias>.

Baule, Giovanni
2012a *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

2012b «Trasfigurazioni di genere. Immagini forti, immagini fragili: il design della comunicazione». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

Baule, Giovanni, Valeria Bucchetti, Elena Caratti, Umberto Tolino, e Marta Reina
2015 «Communication Design for Gender Cultures. A Research and a Design Project for Concrete Actions in the Places of Education». In 11th EAD Conference Proceedings: The Value of Design Research. Sheffield Hallam University. Consultato il 11 giugno 2021, <doi.org/10.7190/ead/2015/169>.

Bender, Emily M., e Batya Friedman
2018 «Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science». *Transactions of the Association for Computational Linguistics 6* (dicembre): 587–604. Consultato il 17 ottobre 2021, <doi.org/10.1162/tacl_a_00041>.

Benenson, Rodrigo, Stefan Popov, e Vittorio Ferrari
2019 «Large-scale interactive object segmentation with human annotators». Aprile. Consultato il 11 gennaio 2022, <arxiv.org/abs/1903.10830>.

Bennato, Davide
2020 «Se (anche) l'algoritmo è sessista: ecco perché Instagram preferisce la pelle femminile nuda». *Network Digital 360*. Consultato il 26 giugno 2021, <agendadigitale.eu/cultura-digitale/se-anche-lalgoritmo-e-sessista-ecco-perche-instagram-preferisce-la-pelle-femminile-nuda/>.

Berger, John
2018 *Questione di sguardi: sette inviti al vedere fra storia dell'arte e quotidianità*. Milano: Il Saggiatore.

2008 *Ways of Seeing*. Penguin on design. London: Penguin.

Best, Mitra, Anand Rao, e Jennifer Lendler
2021 «Understanding algorithmic bias and how to build trust in AI». *PWC, AI and Analytics*. Consultato il 27 giugno 2021, <pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html>.

Biemmi, Irene
2017 *Educazione sessista: Stereotipi di genere nei libri delle elementari.* Rosenberg & Sellier, <doi.org/10.4000/books.res.4626>.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, e Adam Kalai
2016 «Man is to Computer Programmer as Woman is to Homemaker?

Debiasing Word Embeddings». Consultato il 31 luglio 2021, <arxiv.org/abs/1607.06520>.

Bourdieu, Pierre

2014 *Il dominio maschile*. (*La domination masculine*, Seuil, Paris, 1998). Trad. it. Alessandro Serra. 3a ed. Milano: Feltrinelli.

2018 *Un'arte media: saggio sugli usi sociali della fotografia*. Trad. it. Milly Buonanno (a cura di). 3a ed. Milano: Meltemi.

Bucchetti, Valeria

2012 «Modelli quotidiani, stereotipi diffusi». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

2015 *Design e dimensione di genere: un campo di ricerca e riflessione tra culture del progetto e culture di genere*. Design della comunicazione. Snodi 03. Milano: FrancoAngeli.

2017 «Le donne di Lora Lamm». In *Angelica e Bradamante: le donne del design*, 207–24. Biblioteca di architettura 18. Padova: Il poligrafo.

2021 *Cattive immagini. Design della comunicazione, grammatiche e parità di genere*. Design della comunicazione. Snodi 09. Milano: FrancoAngeli.

Bucchetti, Valeria, e Francesca Casnati

2019 «The Contribution of Communication Design to Encourage Gender Equality». In *Designing Sustainability for All: Proceedings of the 3*. LeNS World Distributed Conference, Milano, Mexico City, Beijing, Bangalore, Curitiba, Cape Town, 3-5 April 2019. Milano: Edizioni POLI.design. <re.public.polimi.it/handle/11311/1130363#.YP8UpJMza3I>.

Buolamwini, Joy

2018 *AI, Ain't I A Woman?* Video youtube. Consultato il 17 ottobre 2021, <youtube.com/watch?v=QxuyfWoVV98>.

Buolamwini, Joy, e Timnit Gebru

2018 «Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification». *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research*, 81: 77-91. Consultato il 28 luglio 2021, <proceedings.mlr.press/v81/buolamwini18a.html>.

Burns, Kaylee, Lisa Anne Hendricks, Kate Saenko, Trevor Darrell, e Anna Rohrbach
2019 «Women also Snowboard: Overcoming Bias in Captioning Models». Consultato il 30 luglio 2021, <arxiv.org/abs/1803.09797>.

Caplan, Robyn, Joan Donovan, Lauren Hanson, e Jeanna Matthews
2018 «Algorithmic Accountability: A Primer». *Data&Society*. Consultato il 28 luglio 2021, <datasociety.net/library/algorithmic-accountability-a-primer/>.

Caratti, Elena
2012 «Progettare tra gli stereotipi. Percorsi di ricerca e artefatti per la comunicazione di genere». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

2015 *Rimediazioni gender-sensitive: contributi e progetti per la formazione di un immaginario consapevole*. Design della comunicazione. Snodi 4. Milano: FrancoAngeli.

Carpanini, Clara
2010 «Percorsi di gender performance: fotografia tra arte e moda dagli anni Novanta ad oggi». Application/pdf. Alma Mater Studiorum - Università di Bologna. Consultato il 15 novembre 2021, <amsdottorato.unibo.it/3127>.

Chmielinski, Kasia S, Sarah Newman, Matt Taylor, Josh Joseph, Kemi Thomas, Jessica Yurkofsky, e Yue Chelsea Qiu
2020 «The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence». Consultato il 11 luglio 2021, <securedata.lol/camera_ready/26.pdf>.

Collet, Clementine, e Sarah Dillon
2019 «AI and Gender. Four Proposals for Future Research». University of Cambridge. Consultato il 23 ottobre 2021, <lcfi.ac.uk/media/uploads/files/AI_and_Gender___4_Proposals_for_Future_Research.pdf>.

COMEST World Commission on the Ethics of Scientific Knowledge and Technology
2019 «Preliminary Study on the Ethics of Artificial Intelligence». *UNESCO Digital Library*. Consultato il 26 giugno 2021, <unesdoc.unesco.org/ark:/48223/pf0000367823.locale=en>.

Condliffe, Jamie
2021 «AI Learns Sexism Just by Studying Photographs». *MIT Technology Review*. Consultato il 21 agosto 2021, <technologyreview.com/2017/08/21/149585/ai-learns-sexism-just-by-studying-photographs/>.

Costanza-Chock, Sasha
2018 «Design Justice, A.I., and Escape from the Matrix of Domination». *Journal of Design and Science*, luglio. Consultato il 11 giugno 2021, <doi.org/10.21428/96c8d426>.

Council of Europe Gender Equality Strategy
2016 «Combating Gender Stereotyping and Sexism in the Media». Consultato il 11 giugno 2021, <coe.int/en/web/genderequality/gender-stereotypes-and-sexism>.

Courtland, Rachel
2018 «Bias Detectives: The Researchers Striving to Make Algorithms Fair». *Nature 558* (7710): 357–60. Consultato il 15 novembre 2021, <doi.org/10.1038/d41586-018-05469-3>.

Courtney, Alice E., e Sarah Wernick Lockeretz
1971 «A Woman's Place: An Analysis of the Roles Portrayed by Women in Magazine Advertisements». *Journal of Marketing Research* 8 (1): 92–95. Consultato il 2 ottobre 2021, <doi.org/10.1177/002224377100800114>.

Crawford, Kate
2013 «The Hidden Biases in Big Data». *Harvard Business Review*, Technology. Consultato il 26 giugno 2021, <hbr.org/2013/04/the-hidden-biases-in-big-data>.

Crawford, Kate, e Ryan Calo
2016 «There Is a Blind Spot in AI Research». *Nature 538* (7625): 311–13. Consultato il 26 giugno 2021, <doi.org/10.1038/538311a>.

Crawford, Kate, e Vladan Joler
2018 *Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor*, Data and Planetary Resources. Infografica. Consultato il 26 giugno 2021, <anatomyof.ai/>.

Crawford, Kate, e Trevor Paglen
2019 «Excavating AI: The Politics of Training Sets for Machine Learning». *The AI Now Institute*, NYU. <excavating.ai>.

Crenshaw, Kimberle
1991 «Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color». *Stanford Law Review 43* (6): 1241. Consultato il 28 giugno 2021, <doi.org/10.2307/1229039>.

D'Amico, Marilisa
2020 *Una parità ambigua: costituzione e diritti delle donne*. 1ª ed. Saggi 135. Milano: Raffaello Cortina editore.

Data2X
2019 «Big Data, Big Impact? Towards Gender-Sensitive Data Systems». *Data2X*. Consultato il 3 settembre 2021, <data2x.org/resource-center/big-data-report/>.

Datta, Amit, Michael Carl Tschantz, e Anupam Datta
2015 «Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination». Consultato il 10 ottobre 2021, <arxiv.org/abs/1408.6491>.

Beauvoir, Simone de
2016 *Il secondo sesso*. 2ª ed (1ª ed. 1949 *Le Deuxième Sexe*, Parigi). Tr. it Roberto Cantini e Mario Andreose. Milano: Il saggiatore.

De Piccoli, Norma, e Chiara Rollero, a c. di
2018 «Sui Generi: identità e stereotipi in evoluzione?» In *CIRSDe – Centro Interdisciplinare di Ricerche e Studi delle Donne e di Genere*, 3:280. Convegni. Università degli Studi di Torino. Consultato il 11 giugno 2021. <collane.unito.it/oa/items/show/22#?c=0&m=0&s=0&cv=0>.

De Vries, Patricia, e Willem Schinkel
2019 «Algorithmic Anxiety: Masks and Camouflage in Artistic Imaginaries of Facial Recognition Algorithms». *Big Data & Society* 6 (1). Consultato il 11 giugno 2021, <doi.org/10.1177/2053951719851532>.

Decataldo, Alessandra, e Elisabetta Ruspini
2014 *La ricerca di genere*. Studi superiori 926. Roma: Carocci editore.

Deloitte Insights, a c. di
2020 «The state of AI in the Enterprise, 2nd Edition», *Deloitte Insights*, n. 2: 28. Consultato il 26 giugno 2021, <www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey-2018.html>.

Deni, Michela, e Dario Mangano
2020 «Quando è design». Ocula 21 (ottobre). Consultato il 11 giugno 2021, <doi.org/10.12977/ocula2020-38>.

Diakopoulos, Nicholas
2012 «Understanding bias in computational news media». *NiemanLab, Aggregation & Discovery*. Consultato il 26 giugno 2021, <niemanlab.org/2012/12/nick-diakopoulos-understanding-bias-in-computational-news-media/>.

Diekman, Amanda B., e Alice H. Eagly
2000 «Stereotypes as Dynamic Constructs: Women and Men of the Past, Present, and Future». *Personality and Social Psychology Bulletin* 26 (10): 1171–88. Consultato il 11 giugno 2021, <doi.org/10.1177/0146167200262001>.

D'Ignazio, Catherine, e Lauren F. Klein
2016 «Feminist Data Visualization». In *Workshop on Visualization for the Digital Humanities* (VIS4DH). Baltimore: IEEE. Consultato il 11 giugno 2021, <kanarinka.com/wp-content/uploads/2015/07/IEEE_Feminist_Data_Visualization.pdf>.

2020 *Data feminism*. Strong ideas. Cambridge, Massachusetts: The MIT Press.

Douglas, Laura
2017 «AI is not just learning our biases; it is amplifying them». *Medium*. Consultato il 26 giugno 2021, <medium.com/@laurahelendouglas/ai-is-not-just-learning-our-biases-it-is-amplifying-them-4d0dee75931d>.

Drucker, Johanna
2011 «Humanities Approaches to Graphical Display». D*igital Humanities Quarterly* 5 (1). Consultato il 26 giugno 2021, <digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.

2014 *Graphesis: Visual Forms of Knowledge Production*. Meta LAB Projects. Cambridge, Massachusetts: Harvard University Press.

Durepos, Gabrielle, Alan McKinlay, e Scott Taylor
2017 «Narrating Histories of Women at Work: Archives, Stories, and the Promise of Feminism». *Business History* 59 (8): 1261–79. Consultato il 17 settembre 2021, <doi.org/10.1080/00076791.2016.1276900>.

Dulbecco, Alessia
2020 «Giochi da maschi, giochi da femmine». *Il Tascabile, Società*, giugno. Consultato il 1 dicembre 2021, <iltascabile.com/societa/giochi-da-maschi-giochi-da-femmine/>.

Durham, Meenakshi Gigi, e Douglas Kellner, a c. di
2006 *Media and Cultural Studies: Keyworks*. Rev. ed. Keyworks in Cultural Studies 2. Malden, MA: Blackwell.

Eisend, Martin
2010 «A Meta-Analysis of Gender Roles in Advertising». *Journal of the Academy of Marketing Science* 38 (4): 418–40. Consultato il 11 novembre 2021, <doi.org/10.1007/s11747-009-0181-x>.

EQUALS, UNU, e UNU-CS
2018 «Taking Stock: Data and Evidence on Gender Equality in Digital Access, Skills and Leadership». Consultato il 11 giugno 2021, <2b37021f-0f4a-4640-8352-0a3c1b7c2aab.filesusr.com/ugd/04bfff_145a18e6425e47a1b90d0440f7476d0f.pdf>.

EQUALS, Mark West, Rebecca Kraut, e Han Ei Chew
2019 «I'd Blush If I Could». *EQUALS Skills Coalition*. *UNESCO*. Consultato il 1 settembre 2021, <en.unesco.org/Id-blush-if-I-could>.

Falcinelli, Riccardo
2011 *Guardare, pensare, progettare: neuroscienze per il design*. Stampa Alternativa.

2020 *Figure: come funzionano le immagini dal Rinascimento a Instagram*. Einaudi. Stile libero extra. Torino.

Feenberg, Andrew
2005 *Critical theory of technology*. Vol. 1. Tailoring Biotechnologies 1. New York: Oxford University Press.

2010 *Between reason and experience: essays in technology and modernity*. Inside technology. Cambridge: The MIT Press.

Felix, Shubham Melvin, Sumer Kumar, e A. Veeramuthu
2018 «A Smart Personal AI Assistant for Visually Impaired People». In *2018 2nd International Conference on Trends in Electronics and Informatics* (ICOEI), 1245–50. Tirunelveli: IEEE. Consultato il 11 giugno 2021, <doi.org/10.1109/ICOEI.2018.8553750>.

Feng, Alice, e Shuyan Wu
2019 «The Myth of the Impartial Machine». *Parametric Press*, *Issue 01 Science + Society Table of Contents*. Consultato il 11 giugno 2021, <parametric.press/issue-01/the-myth-of-the-impartial-machine/>.

Ferree, Myra Marx, e Elaine J. Hall
1990 «Visual Images of American Society: Gender and Race in Introductory Sociology Textbooks». *Gender & Society* 4 (4): 500–533. Consultato il 15 settembre 2021, <doi.org/10.1177/089124390004004005>.

Fiedler, Klaus, Claude Messner, e Matthias Bluemke
2006 «Unresolved Problems with the "I", the "A", and the "T": A Logical and Psychometric Critique of the Implicit Association Test (IAT)». *European Review of Social Psychology* 17 (1): 74–147. Consultato il 11 giugno 2021, <doi.org/10.1080/10463280600681248>.

Finn, Ed
2018 *Che cosa vogliono gli algoritmi: l'immaginazione nell'era dei computer*. 2ª ed (1ª ed. 2017 *What Algorithms Want: Imagination in the Age of Computing*. Mit Pr). Tr. it. Daniele A. Gewurz. Torino: Einaudi.

Floridi, Luciano
2017 «The Logic of Design as a Conceptual Logic of Information». *Minds and Machines* 27 (3): 495–519. Consultato il 12 giugno 2021, <doi.org/10.1007/s11023-017-9438-1>.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al.
2018 «AI 4 People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations». *Minds and Machines* 28 (4): 689–707. Consultato il 12 giugno 2021, <doi.org/10.1007/s11023-018-9482-5>.

Frascara, Jorge
2005 *Communication Design: Principles, Methods and Practice*. London: Allworth Press.

Fujiyoshi, Hironobu, Tsubasa Hirakawa, e Takayoshi Yamashita
2019 «Deep Learning-Based Image Recognition for Autonomous Driving». IATSS Research 43 (4): 244–52. Consultato il 26 giugno 2021, <doi.org/10.1016/j.iatssr.2019.11.008>.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, e Kate Crawford
2020 «Datasheets for Datasets». Consultato il 12 giugno 2021, <arxiv.org/abs/1803.09010>.

Gege, Li
2021 «Discrimination is still a problem in STEM». *New Scientist, HUMANS*, maggio. Consultato il 21 luglio 2021, <newscientist.com/article/mg25033331-400-discrimination-is-still-a-problem-in-stem/>.

Gershgorn, Dave
2017 «The Data That Transformed AI Research—and Possibly the World». *Quartz, It's not about the algorithm*. Consultato il 4 settembre 2021, <qz.com/1034972/>.

Gitelman, Lisa, a c. di
2013 *«Raw data» is an oxymoron*. Infrastructures series. Cambridge; London: The MIT Press.

Goffman, Erving.
1969 *La vita quotidiana come rappresentazione*. 2ª ed (1ª ed. 1959 *The Presentation of Self in Everyday Life*). Tr. it Margherita Ciacci. Bologna: Il mulino.

1979 *Gender advertisements*. 1ª ed. 1st Harper colophon ed. New York: Harper & Row.

2015 *Rappresentazioni di genere*. 2ª ed. Tr. it Vanni Codeluppi e Angelo Romeo (a cura di). Milano: Mimesis.

Golden, Amanda, e Cassandra Laity
2018 «Feminist Modernist Digital Humanities». *Feminist Modernist Studies* 1 (3): 205–10. Consultato il 11 giugno 2021, <doi.org/10.1080/24692921.2018.1503786>.

Graff, Kaitlin A., Sarah K. Murnen, e Anna K. Krause
2013 «Low-Cut Shirts and High-Heeled Shoes: Increased Sexualization Across

Time in Magazine Depictions of Girls». Sex Roles 69 (11–12): 571–82. Consultato il 11 giugno 2021, <doi.org/10.1007/s11199-013-0321-0>.

Grau, Stacy Landreth, e Yorgos C. Zotos
2016 «Gender Stereotypes in Advertising: A Review of Current Research». *International Journal of Advertising* 35 (5): 761–70. Consultato il 11 giugno 2021, <doi.org/10.1080/02650487.2016.1203556>.

Greenwald, Anthony G., Debbie E. McGhee, e Jordan L. K. Schwartz
1998 «Measuring Individual Differences in Implicit Cognition: The Implicit Association Test». *Journal of Personality and Social Psychology* 74 (6): 1464–80. Consultato il 17 giugno 2021, <semanticscholar.org/paper/Measuring-individual-differences-in-implicit-the-Greenwald-McGhee/10cc2d53ff8349d3432b8f822d58e4ddee3d475e>.

Guszcza, James, Michelle A Lee, Beena Ammanath, e Dave Kuder
2020 «Human Values in the Loop: Design Principles for Ethical AI», *Deloitte Insights*, 28. Consultato il 11 giugno 2021. <deloitte.com/us/en/insights/focus/cognitive-technologies/design-principles-ethical-artificial-intelligence.html>.

Halliwell, Emma, e Helga Dittmar
2004 «Does Size Matter? The Impact of Model's Body Size on Women's Body-Focused Anxiety and Advertising Effectiveness». *Journal of Social and Clinical Psychology* 23 (1): 104–22. Consultato il 11 novembre 2021, <doi.org/10.1521/jscp.23.1.104.26989>.

Hammond, Kristian
2016 «5 unexpected sources of bias in artificial intelligence». *Tech Crunch*. Consultato il 26 giugno 2021, <techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/>.

Hao, Karen
2019 «AI's White Guy Problem Isn't Going Away». *Technology Review, Artificial intelligence/Machine learning*. Consultato il 17 settembre 2021, <technologyreview.com/2019/04/17/136072/ais-white-guy-problem-isnt-going-away/>.

2021 «An AI Saw a Cropped Photo of AOC. It Autocompleted Her Wearing a Bikini». *Technology Review*, Artificial intelligence/Machine learning. Consultato il 3 settembre 2021, <technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-%20sexist/>.

Heilweil, Rebecca
2020 «Why Algorithms Can Be Racist and Sexist». *Vox, Open Sourced*. Consultato il 27 giugno 2021, <vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency>.

Hentschel, Tanja, Madeline E. Heilman, e Claudia V. Peus
2019 «The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves». *Frontiers in Psychology* 10 (gennaio): 11. Consultato il 11 giugno 2021, <doi.org/10.3389/fpsyg.2019.00011>.

Hesiodus, e Hesiodus
1984 *Teogonia*. Graziano Arrighetti (A cura di). Biblioteca universale Rizzoli I classici della BUR 468. Milano: Rizzoli.

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, e Kasia Chmielinski
2018 «The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards». Consultato il 27 giugno 2021, <arxiv.org/abs/1805.03677>.

Huyskes, Diletta
2020 «AI, Ain't I a Woman? Seeking a New Feminist Ethics of Technology, Between Algorithmic Decision-Making Processes and the European Legislation». Faculty of Humanities University of Leiden. Consultato il 12 giugno 2021, <hdl.handle.net/1887/136660>.

Ideo, a c. di
n.d. «How Can We Use AI to Make Things Better for Humans?» Consultato 11 luglio 2021, <ideo.com/question/how-can-we-use-ai-to-make-things-better-for-humans>.

Il Post
2021 *Questioni di un certo genere: le identità sessuali*, i diritti, le parole da usare : una guida per saperne di più e parlarne meglio. A cura di Arianna Cavallo, Ludovica Lugli, e Massimo Prearo. Vol. 2. *Cose spiegate bene*. Milano: Iperborea.

Isaacson, Lauren
2018 *Boosting Your Bias Immunity*. Video Youtube. Consultato il 26 luglio 2021, <xd.adobe.com/ideas/principles/emerging-technology/removing-ai-bias-pt-1-people-problem/>.

Isidoro Re, Alessandro
2020 «Tech gender bias. Discriminazioni e pregiudizi di genere tra AI e algoritmi». *Singola, Tecnologia*, dicembre. Consultato il 26 giugno 2021, <singola.net/tecnologia/tech-gender-bias-intervista-diletta-huyskes>.

ISTAT Istituto Nazionale di Statistica
2018 *Gli stereotipi sui ruoli di genere*. Consultato il 11 giugno 2021, <istat.it/it/files/2019/11/Report-stereotipi-di-genere.pdf>.

2019 *I tempi della vita quotidiana lavoro, conciliazione, parità di genere e benessere soggettivo.* Consultato il 26 giugno 2021, <stat.it/it/files//2019/05/ebook-I-tempi-della-vita-quotidiana.pdf>.

Jia, Sen, Thomas Lansdall-Welfare, e Nello Cristianini
2015 «Measuring Gender Bias in News Images». In *Proceedings of the 24th International Conference on World Wide Web*, 893–98. Firenze: ACM Press. Consultato il 17 settembre 2021, <doi.org/10.1145/2740908.2742007>.

Jochim, Beth
2021 «Designing AI: The Feminist Way». *Libre AI, AI*. Consultato il 23 settembre 2021, <libreai.com/designing-ai-the-feminist-way/>.

Kabel, Peter, Annika Schultz, e Tom-L. Sager
2021 «Design Influences Society. Conversation on the Era Where Artificial Intelligence and Design Overlap». *Slanted*, maggio 2021.

Kang, Mee-Eun
1997 «The Portrayal of Women's Images in Magazine Advertisements: Goffman's Gender Analysis Revisited». In *Sex-Roles*, di Phyllis A. Katz. Vol. 37. Numbers 9/10. Plenum Press. <link.springer.com/content/pdf/10.1007/BF02936350.pdf>.

Kantayya, Shalini
2020 *Coded Bias*. Documentario. Netflix. Consultato il 11 giugno 2021.

Kay, Matthew, Cynthia Matuszek, e Sean A. Munson
2015 «Unequal Representation and Gender Stereotypes in Image Search Results for Occupations». In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3819–28. Seoul: ACM Press. Consultato il 11 giugno 2021, <doi.org/10.1145/2702123.2702520>.

Kelley, Patrick Gage, Joanna Bresee, Lorrie Faith Cranor, e Robert W. Reeder
2009 «A "Nutrition Label" for Privacy». In *Proceedings of the 5th Symposium on Usable Privacy and Security* - SOUPS '09, 1. Mountain View, California: ACM Press. Consultato il 11 giugno 2021, <doi.org/10.1145/1572532.1572538>.

Kelley, Patrick Gage, Lucian Cesca, Joanna Bresee, e Lorrie Faith Cranor
2010 «Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach». In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*. Atlanta, Georgia, USA: ACM Press. Consultato il 11 giugno 2021, <doi.org/10.1145/1753326.1753561>.

Kimmel, Micheal
2015 *Perché la parità di genere è un bene per tutti — uomini inclusi*. Video Youtube. *TEDWomen*. Consultato il 11 giugno 2021, <ted.com/talks/michael_kimmel_why_gender_equality_is_good_for_everyone_men_included>.

Kitchin, Rob
2014 «Big Data, New Epistemologies and Paradigm Shifts». *Big Data & Society 1* (1). Consultato il 26 giugno 2021, <doi.org/10.1177/2053951714528481>.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, e Cass R Sunstein
2018 «Discrimination in the Age of Algorithms». *Journal of Legal Analysis* 10 (dicembre): 113–74. Consultato il 17 settembre 2021, <doi.org/10.1093/jla/laz001>.

Knight, Will
2017 «Biased Algorithms Are Everywhere, and No One Seems to Care. The big companies developing them show no interest in fixing the problem». *Artificial intelligence/Machine learning, MIT Technology Review*. Consultato il 26 giugno 2021, <technologyreview.com/2017/07/12/150510/biased-algorithms-are-everywhere-and-no-one-seems-to-care>.

Krasin, Ivan, Tom Duerig, Jordi Pont-Tuset, Shahab Kamali, Vittorio Ferrari, Alina Kuznetsova, Hassan Rom, et al.
2017 «OpenImages: A public dataset for large-scale multi-label and multi-class image classification, 2017.» *Open Images Dataset V6*. Consultato il 18 gennaio 2022, <storage.googleapis.com/openimages/web/index.html>.

Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin,

Jordi Pont-Tuset, Shahab Kamali, et al.
2020 «The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale». *International Journal of Computer Vision* 128 (7): 1956–81. Consultato il 11 gennaio 2022, <doi.org/10.1007/s11263-020-01316-z>.

Lages, Cátia Cecília Delgado
2013 «The Design of Nutrition Labels». Escola das Artes da Universidade Católica Portuguesa Mestrado em Gestão de Indústrias Criativas. Consultato il 11 giugno 2021, <hdl.handle.net/10400.14/15775>.

Lambrecht, Anja, e Catherine Tucker
2019 «Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads». *Management Science* 65 (7): 2966–81. Consultato il 11 settembre 2021, <doi.org/10.1287/mnsc.2018.3093>.

Leavy, Susan
2018 «Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning». In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. Gothenburg Sweden: ACM. Consultato il 11 giugno 2021, <doi.org/10.1145/3195570.3195580>.

Leonard, Kim
2021 «What is the Male Gaze? Definition and Examples in Film». *Studio Binder, Theory*. Consultato il 22 giugno 2021, <studiobinder.com/blog/what-is-the-male-gaze-definition/>.

Lessig, Lawrence, e Lawrence Lessig
2006 *Code. Version 2.0.* New York: Basic Books.

Losh, Elizabeth M., e Jacqueline Wernimont, a c. di
2018 *Bodies of Information: Intersectional Feminism and Digital Humanities*. Debates in the Digital Humanities. Minneapolis: University of Minnesota Press.

Lower, Jamie
2018 «Style Speaks: Clothing Judgments, Gender Stereotypes, and Expectancy Violations of Professional Women». *Electronic Theses and Dissertation*s, 2004-2019, University of Central Florida. Consultato il 11 giugno 2021, <stars.library.ucf.edu/etd/5785>.

Lukas, Scott A
2002 «The Gender Ads Project». Consultato il 3 dicembre 2021, <genderads.com>.

Macnish, Kevin
2012 «Unblinking Eyes: The Ethics of Automating Surveillance». *Ethics and Information Technology* 14 (2): 151–67. Consultato il 26 giugno 2021, <doi.org/10.1007/s10676-012-9291-0>.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, e Aram Galstyan
2019 «A Survey on Bias and Fairness in Machine Learning». Consultato il 11 giugno 2021, <arxiv.org/abs/1908.09635>.

Martin, Ashley
2018 «What Happens When We Give Everything a Gender». *Behavioral Scientist*, luglio. Consultato il 2 dicembre 2021, <behavioralscientist.org/what-happens-when-we-give-everything-a-gender/>.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, e Timnit Gebru
2019 «Model Cards for Model Reporting». *Proceedings of the Conference on Fairness, Accountability, and Transparency,* gennaio, 220–29. Consultato il 11 giugno 2021, <doi.org/10.1145/3287560.3287596>.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, e Luciano Floridi
2016 «The Ethics of Algorithms: Mapping the Debate». *Big Data & Society 3* (2). Consultato il 26 giugno 2021, <doi.org/10.1177/2053951716679679>.

Mittelstadt, Brent, Chris Russell, e Sandra Wachter
2019 «Explaining Explanations in AI». In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 279–88. Atlanta, Georgia, USA: ACM Press. Consultato il 25 giugno 2021, <doi.org/10.1145/3287560.3287574>.

Moere, Andrew Vande, e Helen Purchase
2011. «On the Role of Design in Information Visualization». *Information Visualization 10* (4): 356–71. Consultato il 11 giugno 2021, <doi.org/10.1177/1473871611415996>.

Mondzain, Marie-José
2017 *L'immagine che uccide*. 2ª ed. (1ª ed. *L'image peut-elle tuer?*, 2015). Tr. it. Eleonora Montagné. Edb, Bologna.

Morley, Jessica, Luciano Floridi, Libby Kinsey, e Anat Elhalal
2020 «From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices». *Science and Engineering Ethics 26* (4): 2141–68. Consultato il 26 giugno 2021, <doi.org/10.1007/s11948-019-00165-5>.

Mostardini, Massimiliano, Catherine D'Ignazio, Shalini Kantayya, e Ivana Bartoletti
2021 *Intervista di Diletta Huyskes*. Video. Consultato il 24 giugno 2021, <milanodigitalweek.com/tech-gender-bias>.

Mulvey, Laura
1975 « Visual Pleasure and Narrative Cinema». 2ª ed. Rachel Rose e Mark Lewis (a cura di). London: Afterall Books.

Muth, Lisa Charlotte
2018 «An Alternative to Pink & Blue: Colors for Gender Data». *Datawrapper, Data Vis Dos & Don'ts*, luglio. Consultato il 22 luglio 2021, <blog.datawrapper.de/gendercolor/>.

Ntoutsi, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria Esther Vidal, Salvatore Ruggieri, et al.
2020 «Bias in Data driven Artificial Intelligence Systems—An Introductory Survey». *WIREs Data Mining and Knowledge Discovery 10* (3). Consultato il 27 giugno 2021, <doi.org/10.1002/widm.1356>.

O'Neil, Cathy
2016a «How Algorithms Rule Our Working Lives». *The Guardian, The long read*, settembre. Consultato il 26 luglio 2021, <theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives>.

2016b *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. First edition. New York: Crown.

Paoli, Julia
2008 «Deconstructing "Woman"». Western University of Canada. Consultato il 1 settembre 2021, <uwo.ca/visarts/research/2008-09/bon_a_tirer/Julia%20Paoli.html>.

Piet, Nadia
2020 «AI Meets Design Toolkit». *MOBGEN - Accenture Interactive*. Consultato il 11 maggio 2021, <aimeets.design/>.

Plakoyiannaki, Emmanuella, Kalliopi Mathioudaki, Pavlos Dimitratos, e Yorgos Zotos
2008 «Images of Women in Online Advertisements of Global Products: Does Sexism Exist?» *Journal of Business Ethics 83* (1): 101–12. Consultato il 11 novembre 2021, <doi.org/10.1007/s10551-007-9651-6>.

Planned Parenthood, a c. di
2021 «What are gender roles and stereotypes?» *Planned Parenthood*. Consultato il 11 novembre 2021, <plannedparenthood.org/learn/gender-identity/sex-gender-identity/what-are-gender-roles-and-stereotypes>.

Pochettino, Silvia
2020 «Perchè l'Intelligenza Artificiale è (in)consapevolmente sessista». *Ong2zero, AI, data visualisation*. Consultato il 26 luglio 2021, <ong2zero.org/blog/perche-lintelligenza-artificiale-e-inconsapevolmente-sessista/>.

Raji, Inioluwa Deborah, e Jingying Yang
2020 «ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles». Consultato il 11 giugno 2021, <arxiv.org/abs/1912.06166>.

Richards, John, David Piorkowski, Michael Hind, Stephanie Houde, e Aleksandra Mojsilović
2020 «A Methodology for Creating AI FactSheets». Consultato il 11 giugno 2021, <arxiv.org/abs/2006.13796>.

Ricks, Becca, e Mark Surman
2020 «Creating Trustworthy AI a Mozilla White Paper on Challenges and Opportunities in the AI Era». *Mozilla Foundation*. Consultato il 11 giugno 2021, <foundation.mozilla.org/en/insights/trustworthy-ai-whitepaper/>.

Rogers, Richard
2013 *Digital Methods*. London: The MIT Press.

2019 «Tracker Analysis Detection Techniques for Data Journalism Research». In *Doing Digital Methods*, SAGE, 229–46. Los Angeles: SAGE Publications, Inc.

Sarikakis, Katharine
2013 «Media and the Image of Women». In *1st Conference of the Council of Europe Network of National Focal Points on Gender Equality*, 30, Amsterdam. Consultato il 26 giugno 2021, <rm.coe.int/1680590587>.

Scanu, Claudia
2012 «Impressioni maschili su pagine femminili. Lo sguardo maschile nella rappresentazione pubblicitaria della donna». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

Schmieg, Sebastian
2019 «Decisive Mirror». Portfolio. Sebastian Schmieg. Consultato il 23 settembre 2021, <sebastianschmieg.com/decisive-mirror/>.

Schmitt, David P.
2017 «On That Google Memo About Sex Differences. A response to claims psychological sex differences are "incorrect assumptions"». *Psychology Today, SEX, Sexual Personalities* (August). Consultato il 20 gennaio 2022, <psychologytoday.com/intl/blog/sexual-personalities/201708/google-memo-about-sex-differences>.

Schwab, Klaus, Robert Crotti, Thierry Geiger, Vesselina Ratcheva, e World Economic Forum
2019 *Global Gender Gap Report 2020 Insight Report*. Geneva: World Economic Forum.

Schwemmer, Carsten, Carly Knight, Emily D. Bello-Pardo, Stan Oklobdzija, Martijn Schoonvelde, e Jeffrey W. Lockhart
2020 «Diagnosing Gender Bias in Image Recognition Systems». *Socius: Sociological Research for a Dynamic World 6* (gennaio). Consultato il 15 novembre 2021, <doi.org/10.1177/2378023120967171>.

Severi, Alessio
2018 «Tra stereotipi di genere e realtà». *Community LGBTQI, News*. Consultato il 22 luglio 2021, <noimtfftm.altervista.org/index.php/portfolio/item/36-gli-stereotipi-di-genere/36-gli-stereotipi-di-genere?jjj=1629628957670>.

Shankar, Shreya, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, e D. Sculley
2017 «No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World». Consultato il 26 giugno

2021, <arxiv.org/abs/1711.08536>.

Sherriffs, Alex C., e John P. McKEE
1957 «Qualitative Aspects of Beliefs About Men and Women». *Journal of Personality* 25 (4): 451–64. Consultato il 17 settembre 2021, <doi.org/10.1111/j.1467-6494.1957.tb01540.x>.

Silberg, Jake, e James Manyika
2019 «Notes from the AI Frontier: Tackling Bias in AI (and in Humans)». *McKinsey Global Institute*. Consultato il 11 giugno 2021, <mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>.

Silva, Selena, e Martin Kenney
2019 «Algorithms, Platforms, and Ethnic Bias». *Communications of the ACM 62 (11): 37–39*. Consultato il 26 giugno 2021, <doi.org/10.1145/3318157>.

Simonite, Tom
2017 «Machines Taught by Photos Learn a Sexist View of Women». *Wired, Business*, agosto. Consultato il 31 luglio 2021, <wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>.

2018a «When It Comes to Gorillas, Google Photos Remains Blind». *Wired, Business*, gennaio. Consultato il 26 giugno 2021, <wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>.

2018b «AI Is the Future—But Where Are the Women?» *Wired, Business*, agosto. Consultato il 31 giugno 2021, <wired.com/story/artificial-intelligence-researchers-gender-imbalance/>.

2020 «When AI Sees a Man, It Thinks “Official.” A Woman? “Smile”». *Wired, Business*, novembre. Consultato il 1 settembre 2021, <wired.com/story/ai-sees-man-thinks-official-woman-smile/>.

Sinders, Caroline
2016 «Microsoft’s Tay Is an Example of Bad Design. Why Interaction Design Matters, and so Does QA-Ing». *Medium*. Consultato il 26 giugno 2021, <medium.com/@carolinesinders/microsoft-s-tay-is-an-example-of-bad-design-d4e65bb2569f>.

2017 «Feminist Data Set». *University of Denver.* Consultato il 11 giugno 2021, <carolinesinders.com/feminist-data-set/>.

2019 *AI Is More than Math: Using Design to Confront Bias*. Youtube video. #AIGADesignConf. Consultato il 26 giugno 2021, <youtube.com/watch?v=NIi0g94QlvY>.

Spielkamp, Matthias
2017 «Inspecting Algorithms for Bias». *Technology Review*, giugno. Consultato il 11 giugno 2021, <technologyreview.com/2017/06/12/105804/inspecting-algorithms-for-bias/>.

Steed, Ryan, e Aylin Caliskan
2021 «Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases». *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, marzo, 701–13. Consultato il 2 settembre 2021, <doi.org/10.1145/3442188.3445932>.

Stoyanovich, Julia, e Bill Howe
2019 «Nutritional Labels for Data and Models». *IEEE Data Eng. Bull., Computer Science*, 42: 13–23. Consultato il 11 giugno 2021, <sites.computer.org/debull/A19sept/p13.pdf>.

Suresh, Harini, e John V. Guttag
2021 «A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle». Consultato il 11 giugno 2021, <arxiv.org/abs/1901.10002>.

Sydell, Laura
2016 «It Ain't Me, Babe: Researchers Find Flaws In Police Facial Recognition Technology», *Privacy & Security*. Consultato il 26 giugno 2021, <npr.org/sections/alltechconsidered/2016/10/25/499176469/it-aint-me-babe-researchers-find-flaws-in-police-facial-recognition?t=1627327966990>.

TED Talks
2017 *The era of blind faith in big data must end*. Video Youtube. *TED*. Consultato il 27 giugno 2021, <youtube.com/watch?v=2ueHHzRto>.

Tedesco, Lucia
2021 «Etica digitale e discriminazioni di genere, l'intervista a Diletta Huyskes». Consultato il 26 giugno 2021, <techprincess.it/etica-digitale-intervista-diletta-huyskes/>.

Teigland, Julie Linn
2019 «Why we need to solve the issue of gender bias before AI makes it

worse». Consultato il 24 giugno 2021, <ey.com/en_gl/wef/why-we-need-to-solve-the-issue-of-gender-bias-before-ai-makes-it>.

Thelwall, Mike
2018 «Gender Bias in Machine Learning for Sentiment Analysis». *Online Information Review* 42 (3): 343–54. Consultato il 11 giugno 2021, <doi.org/10.1108/OIR-05-2017-0153>.

Tolino, Umberto
2012 «Cataloghi per educare. Riconfigurazione di tipi femminili». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

Turner Lee, Nicol
2019 «Inclusion in Tech: How Diversity Benefits All Americans». *Brookings EDU, Testimony*. Consultato il 25 giugno 2021, <brookings.edu/testimonies/inclusion-in-tech-how-diversity-benefits-all-americans/>.

Van de Walle, Josephine
2017 «Modern Classics: Cindy Sherman - Untitled Film Stills, 1977-1980». *Art Lead, modern classics*. Consultato il 17 settembre 2021, <artlead.net/content/journal/modern-classics-cindy-sherman-untitled-film-stills/>.

Venables, Michelle
2019 «An Overview of Computer Vision». *Toward Data Science*, settembre. Consultato il 5 dicembre 2021, <towardsdatascience.com/an-overview-of-computer-vision-1f75c2ab1b66>.

Villano, Paola
2003 *Pregiudizi e stereotipi*. 1ª ed. Scienze sociali 104. Roma: Carocci.

Vincent, James
2018 «Google 'Fixed' Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech». *The Verge*. Consultato il 26 giugno 2021. <theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>.

Wang, Angelina, Alexander Liu, Ryan Zhang, Anat Kleiman, Leslie Kim, Dora Zhao, Iroha Shirai, Arvind Narayanan, e Olga Russakovsky
2021 «REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets», luglio. Consultato il 26 giugno 2021. <arxiv.org/abs/2004.07999>.

West, Sarah Myers, Kate Crawford, e Meredith Whittaker
2019 «Discriminating Systems: Gender, Race and Power in AI». *AI Now Institute*. Consultato il 17 settembre 2021, <ainowinstitute.org/discriminatingsystems.html>.

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Kaziunas, Mathur, Myers West, Richardson, Schultz, e Schwartz
2018 «AI Now Report 2018». *AI Now Institute*. Consultato il 11 giugno 2021, <ainowinstitute.org/AI_Now_2018_Report.pdf>.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, e Stephen Cave
2019 «The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions». In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200. Honolulu HI USA: ACM. Consultato il 11 giugno 2021, <doi.org/10.1145/3306618.3314289>.

Wikipedia
2020 «Paradosso di Simpson». Consultato il 26 giugno 2021, <it.wikipedia.org/wiki/Paradosso_di_Simpson>.

Wood, Julia T
1994 «Gendered Media: The Influence of Media on Views of Gender». Consultato il 11 giugno 2021, <www1.udel.edu/comm245/readings/GenderedMedia.pdf>.

World Economic Forum
2021 «Global Gender Gap Report 2021». Consultato il 29 luglio 2021, <reports.weforum.org/globalgender-gap-report-2021/dataexplorer>.

Yang, Ke, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, H. V. Jagadish, e Gerome Miklau
2018 «A Nutritional Label for Rankings». *Proceedings of the 2018 International Conference on Management of Data*, maggio, 1773–76. Consultato il 11 giugno 2021, <doi.org/10.1145/3183713.3193568>.

Yatskar, Mark, Luke Zettlemoyer, e Ali Farhadi
2016 «Situation Recognition: Visual Semantic Role Labeling for Image Understanding». In *2016 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 5534–42. Las Vegas, NV, USA: IEEE. Consultato il 18 gennaio 2022, <doi.org/10.1109/CVPR.2016.597>.

Zer-Aviv, Mushon
2018 «The Normalizing Machine». Consultato il 21 settembre 2021, <mushon.com/tnm/>.

Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez, e Kai-Wei Chang
2017 «Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints». Consultato il 30 luglio 2021, <arxiv.org/abs/1707.09457>.

Zingale, Salvatore
2012 «Immobili visioni. Domande intorno alla persistenza dello stereotipo». In *Anticorpi comunicativi. Progettare per la Comunicazione di genere*, di Giovanni Baule, Valeria Bucchetti (a cura di). Milano: FrancoAngeli.

Zou, James, e Londa Schiebinger
2018a «AI Can Be Sexist and Racist — It's Time to Make It Fair». *Springer Nature volume 559*: 324–26. Consultato il 29 luglio 2021, <doi.org/10.1038/d41586-018-05707-8>.

2018b «Design AI so That It's Fair». *Springer Nature Limited volume 559 (luglio): 3*. Consultato il 11 giugno 2021, <nature.com/articles/d41586-018-05707-8>.

Indice delle figure

Gli algoritmi sono ai più incomprensibili e in tal senso trattati come universalmente veritieri. Per esistere si nutrono di dati che sono però scelti, selezionati e filtrati da esseri umani e infine da questi ultimi implementati. Nel processo si annidano molteplici bias, tra cui spiccano per numero i gender bias. La tesi vuole indagare come la dimensione del progetto di comunicazione possa supportare la ricerca e lo sviluppo di intelligenze artificiali più eque ed inclusive. Avvalendosi della lente di decostruzione del design e degli studi di genere, il fine ultimo è fornire uno strumento che possa individuare la presenza di gender bias nei contenuti visivi di allenamento delle intelligenze artificiali.

Beatrice Bazzan

Laurea magistrale in Design della Comunicazione
Scuola del Design, Politecnico di Milano
A.A. 2019/2020