



UNIVERSITÀ DEGLI STUDI DI MILANO

*Facoltà di Giurisprudenza*

*Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale*

---

**IA E DISCRIMINAZIONI DI GENERE  
PROSPETTIVE FUTURE  
PER UNA REGOLAMENTAZIONE INCLUSIVA**

*a cura di*

**MARILISA D'AMICO e NICCOLÒ PANIGADA**



**G. Giappichelli Editore**



UNIVERSITÀ DEGLI STUDI DI MILANO

*Facoltà di Giurisprudenza*

*Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale*

---

Studi di diritto pubblico

125

La Collana “Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale” dell’Università degli Studi di Milano raccoglie monografie e altri risultati inediti di ricerche, individuali e collettive, di studiosi che svolgono attività di studio e ricerca nel Dipartimento.

Essa comprende Studi di Diritto costituzionale, di Diritto amministrativo, di Diritto internazionale ed europeo, di Diritto processuale civile, di Diritto comparato, di Storia del diritto, di Politica economica.

La qualità scientifica delle pubblicazioni è assicurata da una procedura di c.d. double blind peer review ad opera di revisori esterni.

IA E DISCRIMINAZIONI DI GENERE  
PROSPETTIVE FUTURE  
PER UNA REGOLAMENTAZIONE INCLUSIVA

*a cura di*

MARILISA D'AMICO e NICCOLÒ PANIGADA



G. Giappichelli Editore

© Copyright 2025 - G. GIAPPICHELLI EDITORE - TORINO

VIA PO, 21 - TEL. 011-81.53.111

<http://www.giappichelli.it>

ISBN/EAN 979-12-211-1825-4

ISBN/EAN 979-12-211-6583-8 (ebook-pdf)

ISBN/EAN 979-12-211-6721-4 (ebook-epub)

*Il volume è edito con risorse del PRIN 2022 CUP G53D23002160006 - Finanziato dall'Unione europea - Next Generation EU Misura M4C2 – Investimento 1.1.*



*Stampa:* Stampatre s.r.l. - Torino

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941, n. 633.

Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali, Corso di Porta Romana 108, 20122 Milano, e-mail [autorizzazioni@clearedi.org](mailto:autorizzazioni@clearedi.org) e sito web [www.clearedi.org](http://www.clearedi.org).

## INDICE

	<i>pag.</i>
<i>Elenco degli Autori e delle Autrici</i>	XV
<i>Note introduttive sul PRIN – AiGeDi (Artificial Intelligence between generating and tackling gender-based discriminations; L'intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere)</i> <i>Marilisa D'Amico</i>	XVII
RELAZIONE INTRODUTTIVA	
SFIDE PER COMBATTERE LE DISCRIMINAZIONI DI GENERE NELL'ERA DELL'IA	
<i>Francisco Balaguer Callejón</i>	1
1. Introduzione	1
2. L'uguaglianza di genere come ultima frontiera del diritto costituzionale	3
3. L'era dell'IA e l'algoritmo come nuovo strumento di trasformazione storica	5
4. La parità di genere nell'era dell'intelligenza artificiale	7
4.1. L'orientamento dei diritti verso un mercato discriminatorio	7
4.2. Parità di genere e riproduzione culturale della discriminazione nell'era digitale	10
5. Conclusioni	13

## PARTE I

DIRITTI COSTITUZIONALI, IDENTITÀ  
E *BIAS* ALGORITMICI: PROSPETTIVE  
GIURIDICHE NELL'ERA DELL'IA

## CAPITOLO I

LE NUOVE FRONTIERE DELLE DISCRIMINAZIONI DI  
GENERE: L'IMPATTO DELL'INTELLIGENZA ARTIFICIALE*Marilisa D'Amico*

17

- |   |    |
|---|----|
| 1. Intelligenza artificiale e discriminazioni di genere                           | 17 |
| 2. La discriminazione algoritmica e i suoi diversi fattori                        | 19 |
| 3. Il linguaggio degli algoritmi e l'impatto discriminatorio                      | 23 |
| 4. Lavoro femminile e IA  | 28 |
| 5. AI Act europeo e diritti delle donne   | 29 |
| 6. Considerazioni sul Trattato del Consiglio d'Europa                             | 30 |
| 7. Una nuova sfida: l'IA americana e il "free speech" algoritmico                 | 32 |
| 8. Postilla: la legge italiana e il riferimento specifico alla "parità di genere" | 33 |

## CAPITOLO II

CONCEPTUAL EVOLUTION AND NORMATIVE TENSION  
OF PERSONAL IDENTITY BETWEEN AI TECHNOLOGIES  
AND BEYOND*Edoardo Carlo Raffiotta*

37

- |  |    |
|--|----|
| 1. Background  | 37 |
| 2. The necessary protection of personal identity in the digital world  | 38 |
| 3. The necessary protection of personal identity in the era of AI  | 44 |
| 3.1. From the virtual to the metaverse, opportunities and criticalities of the<br>new dimensions of personal development | 47 |
| 4. Over-regulation in the EU and the need for fewer, more effective rules  | 52 |
| 5. Between AI and metaverse technology   | 55 |
| 6. A provocation. Conclusions  | 55 |

*pag.*

## CAPITOLO III

L'INTELLIGENZA ARTIFICIALE DA CAUSA A  
STRUMENTO DI CONTRASTO DELLE DISCRIMINAZIONI  
DI GENERE E INTERSEZIONALI: CONSIDERAZIONI  
A SEGUITO DI INDAGINE "DELPHI"*Elettra Stradella, Silvia Cervia, Gabriele Rugani* 57

1. Introduzione: domanda di ricerca e metodologia 57
2. Le discriminazioni di genere e intersezionali derivanti dall'intelligenza artificiale 59
3. Le cause delle discriminazioni algoritmiche 61
4. Gli ambiti in cui si possono verificare gli effetti pregiudizievoli derivanti dall'IA 63
5. Il ruolo degli strumenti giuridici esistenti a livello UE nel prevenire e contrastare le discriminazioni derivanti dall'IA 65
6. Le strategie per addivenire a un'IA che non causi, e addirittura contrasti, discriminazioni di genere e intersezionali, con particolare riferimento al coinvolgimento degli *stakeholders* 69
7. Considerazioni conclusive 71

## CAPITOLO IV

CERTIFICAZIONE DI PROVENIENZA DELLE RISPOSTE PER  
L'USO PROFESSIONALE DEI *LARGE LANGUAGE MODEL**Ernesto Damiani* 73

1. Introduzione 73
2. Provenienza delle risposte LLM 75
  - 2.1. RAG con marcatori verificabili 76
3. Stato dell'arte tecnologico 79
4. Questioni aperte e direzioni future di ricerca 80
5. Conclusioni 81

## CAPITOLO V

I DIRITTI DELLE PERSONE CON DISABILITÀ AL  
COSPETTO DEI SISTEMI DI INTELLIGENZA ARTIFICIALE.  
UN'ANALISI DELL'AI ACT E DELLA LEGGE N. 132/2025*Giuseppe Arconzo, Gaia Patarini* 83

1. Disabilità e sistemi di intelligenza artificiale: opportunità e rischi 83

	<i>pag.</i>
2. Le persone con disabilità come soggetti vulnerabili da tutelare secondo l'AI Act	87
3. ( <i>Segue</i> ): le altre condizioni richieste per l'applicazione del sistema di tutela nell'AI Act	93
4. Il rilievo, nell'AI Act, dell'accessibilità dei sistemi di IA	95
5. Persone con disabilità e intelligenza artificiale nella legge italiana: sfide e potenzialità di uno strumento in divenire ... ma a costo zero	98
6. Conclusioni	102

## CAPITOLO VI

MISOGINIA E *HATE SPEECH ONLINE* NEL DIRITTO DELL'UE

*Ilaria Anrò* 105

1. Introduzione: la lotta alla misoginia attraverso il contrasto all' <i>hate speech online</i> nel diritto dell'Unione europea	105
2. <i>Hate speech online</i> e strumenti di cooperazione giudiziaria penale	109
3. L'inserimento di <i>hate speech</i> e <i>hate crimes</i> tra i "reati europei"	111
4. La risoluzione del Parlamento europeo del 18 gennaio 2024 a favore dell'estensione della lista <i>ex art. 83 TFUE</i>	115
5. Gli ostacoli tuttora esistenti e il cambio di strategia della Commissione	116
6. Alcune riflessioni conclusive	120

## CAPITOLO VII

## SISTEMI DI INTELLIGENZA ARTIFICIALE E VIOLENZA DI GENERE: DUE ITINERARI DI RIFLESSIONE TRA LORO CONVERGENTI

*Irene Pellizzone* 121

1. Considerazioni introduttive sull'impatto dei sistemi di intelligenza artificiale sul fenomeno della violenza di genere: un itinerario a due sensi	121
2. I punti di forza dei Sistemi di intelligenza artificiale: strumenti di contrasto alla violenza di genere (richieste di aiuto ed esame dei fattori di rischio)	125
3. Le donne come vittime predestinate dell'IA: <i>focus</i> sul <i>Deep Fake</i> non consensuale	129
4. Un importante passo avanti: l'intervento normativo attraverso la recente legge n. 132/2025	135
5. Prevenzione della violenza nei confronti delle donne e dell'intelligenza artificiale: un primo bilancio	136

*pag.*

## CAPITOLO VIII

SALUTE DELLE DONNE E REGOLAMENTAZIONE  
DELL'IA: LA *GENDER-SENSITIVE GOVERNANCE* COME  
PARADIGMA DI EGUAGLIANZA SOSTANZIALE*Marta Tomasi*

139

1. Introduzione. L'irrealizzato obiettivo dell'eguaglianza in salute 139
2. Salute delle donne e discriminazioni strutturali 140
3. L'automatizzazione delle diseguaglianze 142
4. La produzione del dato come spazio di diseguaglianza: la ricerca biomedica tra esclusione e *bias* 144
5. La pratica clinica come spazio di diseguaglianza: diagnosi, cura e dimensione strutturale della salute 148
6. La ricerca dell'eguaglianza di genere nello spazio della regolazione 152
7. Soluzioni normative a confronto: equità sostanziale e *gender-sensitive governance* 156

## CAPITOLO IX

I SISTEMI DI IA PER IL CONTROLLO  
DELL'IMMIGRAZIONE: BENEFICI E RISCHI  
PER I DIRITTI UMANI E LA PARITÀ DI GENERE*Cecilia Siccardi*

161

1. Premessa 161
2. *Smart borders* e *governance* digitale dell'immigrazione 162
3. L'impatto dei sistemi di IA sulla condizione delle donne migranti 166
4. L'AI Act e la disciplina dei sistemi per il controllo dell'immigrazione 169
5. Le garanzie previste per i sistemi "ad alto rischio" ed alcune eccezioni 172
6. Alcuni ostacoli all'applicabilità delle garanzie previste dall'AI Act 175
7. I sistemi vietati che vengono in rilievo nel contesto dell'immigrazione 177
8. Riflessioni conclusive 178

## CAPITOLO X

IL POTERE PERFORMATIVO DEL LINGUAGGIO NELL'ERA  
DELL'IA*Nannerel Fiano*

179

1. L'intelligenza artificiale e il suo linguaggio *vis à vis* con la Costituzione 179

	<i>pag.</i>
2. Il funzionamento dei LLM	181
3. L' <i>hate speech</i> , questo (s)conosciuto: dalla filosofia del linguaggio alla tutela costituzionale	183
4. La regolamentazione europea dell'intelligenza artificiale: quale linguaggio?	185
5. La legge n. 132/2025 e i profili linguistici	186
6. Conclusioni: per un linguaggio umano e tecnologico non discriminatorio	187

## CAPITOLO XI

### LA PROTEZIONE DEI DATI PERSONALI AL TEMPO DELL'INTELLIGENZA ARTIFICIALE: IL COORDINAMENTO TRA AI ACT E GDPR

*Pietro Villaschi* 189

1. Introduzione: i dati come linfa vitale della società digitale	189
2. La perdurante attualità di una regolazione che limiti un trattamento “selvaggio” dei dati personali degli utenti della società digitale	190
3. AI ACT e GDPR: punti di contatto e complementarità delle normative europee	192
4. AI ACT e GDPR: possibili problemi di coordinamento	193
5. Tra proposte di integrazione e rischi di sovra-regolazione	195

## CAPITOLO XII

### AI ACT E DISCRIMINAZIONI DI GENERE: TUTELA EXTRA-GIUDIZIALE E GIUDIZIALE A CONFRONTO

*Paolo Gambatesa* 197

1. Introduzione: l'algoritmo in una prospettiva costituzionale quale chiave di lettura per il contrasto delle discriminazioni di genere	197
2. Il contrasto alle discriminazioni di genere nei sistemi vietati e ad alto rischio. Una lettura in filigrana dell'AI Act	201
3. ( <i>Segue</i> ): le previsioni più esplicite della legge n. 132/2025	205
4. Luci e ombre dello strumento extra-giudiziale della valutazione di impatto ai sensi dell'art. 27 dell'AI Act	206
5. Le possibili prospettive di tutela giudiziale in un'ottica di integrazione dei livelli di protezione dei diritti	209
6. Considerazioni conclusive	212

*pag.*

## CAPITOLO XIII

L'UTILIZZO DI SISTEMI DI IDENTIFICAZIONE  
BIOMETRICA REMOTA NEL REGOLAMENTO  
SULL'INTELLIGENZA ARTIFICIALE TRA MECCANISMO DI  
AUTORIZZAZIONE E MECCANISMO DI SORVEGLIANZA*Giulia Agrati*

215

1. Premessa 215
2. Introduzione: il divieto di utilizzo di sistemi di identificazione biometrica remota in "tempo reale" in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità e le rispettive eccezioni 217
3. Il meccanismo di autorizzazione e il ruolo dell'autorità giudiziaria o autorità amministrativa indipendente 219
4. Il meccanismo di controllo e il ruolo dell'autorità di vigilanza del mercato e dell'autorità nazionale per la protezione dei dati 224
5. Conclusioni 226

## CAPITOLO XIV

LE DISCRIMINAZIONI DI GENERE NELL'ACCESSO AL  
LAVORO GENERATE O AMPLIFICATE DAI SISTEMI DI  
INTELLIGENZA ARTIFICIALE: PROFILI COSTITUZIONALI*Niccolò Panigada*

227

1. I sistemi di IA e la discriminazione di genere nell'accesso al lavoro: origine del problema a partire da alcuni casi concreti 227
2. L'ancoraggio costituzionale e sovranazionale 229
3. La legge italiana sull'intelligenza artificiale in relazione al mondo del lavoro 231
4. Rischi e criticità giuridiche ancora presenti 232
5. I rimedi previsti dall'ordinamento 234
6. Conclusioni 235

## CAPITOLO XV

IA GENERATIVA E DIFFUSIONE DEGLI STEREOTIPI  
ONLINE. LA CONCETTUALIZZAZIONE DI UN AGENTE  
CONTRO IL LINGUAGGIO DISCRIMINATORIO*Marta Annamaria Tamborini*

237

1. Premessa 237

	<i>pag.</i>
2. Le nuove frontiere della discriminazione algoritmica	239
3. <i>Bias</i> nel linguaggio, <i>bias</i> nell'algoritmo e diffusione di odio	241
4. L'IA generativa come alleato e non come minaccia	243
5. Il progetto dell'identificazione degli stereotipi e del linguaggio d'odio attraverso l' <i>Agent for Anti Discriminatory Language</i> (AAL)	243
6. Conclusioni	245

## PARTE II

### GOVERNANCE INCLUSIVA DELL'INTELLIGENZA ARTIFICIALE: PRATICHE ISTITUZIONALI, FORENSI E AZIENDALI

#### CAPITOLO XVI

##### INTELLIGENZA ARTIFICIALE, DISCRIMINAZIONI DI GENERE E PUBBLICHE AMMINISTRAZIONI

*Elena Buscemi*

249

#### CAPITOLO XVII

##### INTELLIGENZA ARTIFICIALE E DISCRIMINAZIONI DI GENERE: IL RUOLO DELLE ISTITUZIONI FORENSI NEL GOVERNO DI UN ECOSISTEMA INCLUSIVO

*Antonino La Lumia*

253

1. Premessa: perché parlare di IA, genere e avvocatura	253
2. IA e discriminazioni di genere: forme, casi, meccanismi	254
2.1. Dal pregiudizio sociale al <i>bias</i> algoritmico	254
2.2. Occupazione, selezione del personale e valutazione delle performance	254
3. L' <i>AI Act</i> europeo e la centralità del rischio discriminatorio	255
4. IA, genere e responsabilità delle istituzioni forensi	256
4.1. Dall'etica "dichiarata" alla responsabilità "praticata"	256
4.2. IA e dovere di riservatezza: una questione anche di genere	256
5. Parità di genere e accessibilità economica ai sistemi di IA	257
6. IA come strumento di contrasto delle discriminazioni: dal rischio all'opportunità	258
7. Conclusioni: un'agenda condivisa per un'IA inclusiva	258

	<i>pag.</i>
CAPITOLO XVIII	
INTELLIGENZA ARTIFICIALE RESPONSABILE, UN VANTAGGIO COMPETITIVO: UN FRAMEWORK AZIENDALE PER INNOVAZIONE, ETICA E SOSTENIBILITÀ	
<i>Gabriele Secol, Beatrice Zulianello</i>	261
SEZIONE I	
L'AI ACT IN AZIENDA	
1. Gli impatti delle regolamentazioni europee nelle società di consulenza	261
2. L'iniziativa aziendale	262
SEZIONE II	
INNOVAZIONE RESPONSABILE	
3. La sostenibilità dell'innovazione	263
4. Algoritmi discriminatori	264
5. Processo di gestione dei <i>bias</i> nei sistemi di intelligenza artificiale	266
6. Sensibilità degli strumenti	267
SEZIONE III	
IMPEGNO DI DELOITTE	
7. Promozione dell'innovazione per il sociale	269
CAPITOLO XIX	
INTELLIGENZA ARTIFICIALE TRA ROTTURA E TRANSIZIONE: DAL "RED FLAG ACT" ALLO <i>HUMAN OVERSIGHT</i>	
<i>Giuseppe Vaciago</i>	271
1. Introduzione: la paura "sbagliata" e il rischio di una nuova grande rottura	271
SEZIONE I	
GOVERNARE SENZA FRENARE	
2. L'AI Act: ambizione, cornice di tutela e rischio di <i>over</i> -regolazione	273
3. Lo <i>Human Oversight Officer</i> : profilo e ruolo operativo	273
SEZIONE II	
POTERI PRIVATI E DIRITTO PUBBLICO	
4. Le piattaforme come nuovi regolatori privati	275

	<i>pag.</i>
5. Trasparenza, <i>auditing</i> e tutela effettiva degli utenti	276
SEZIONE III	
UN LESSICO PER IL FUTURO	
6. Dall'informatica giuridica al diritto dell'informatica	277
7. Proposte operative per i corpi intermedi e gli Ordini professionali	278
8. Conclusioni	279
CAPITOLO XX	
L'INTELLIGENZA ARTIFICIALE IN A2A	
<i>Patrick Oungre</i>	281
1. Introduzione: <i>overview</i> sull'utilizzo e sviluppo di sistemi di intelligenza artificiale in A2A	281
2. I pillar di A2A e i progetti	281
2.1. L'obiettivo	281
2.2. Agenti IA a supporto dei team	282
2.3. Tecnologie agentiche a supporto dei processi	282
2.4. L'IA fisica	283
2.4.1. Progetto Prometeo	283
2.4.2. Progetto Pandora: manutenzione predittiva della rete elettrica	284
2.4.3. Progetto Ottimizzazione calore erogato impianti di teleriscaldamento	285
CAPITOLO XXI	
INTELLIGENZA ARTIFICIALE E IMPRESA: TRA INNOVAZIONE TECNOLOGICA E RESPONSABILITÀ CULTURALE	
<i>Alessandro Renna</i>	287
1. Premessa	287
2. Il vero problema non è la tecnologia	287
3. L'IA come leva competitiva delle organizzazioni: la nostra esperienza	288
4. L'IA come abilitatore di una nuova offerta tecnologica in ambito <i>legal</i>	289
5. Questioni interpretative aperte: il concetto di prevalenza	290
6. Conclusioni: un approccio europeo	291

CAPITOLO I

LE NUOVE FRONTIERE DELLE DISCRIMINAZIONI DI  
GENERE: L'IMPATTO DELL'INTELLIGENZA ARTIFICIALE

*Marilisa D'Amico*

SOMMARIO: 1. Intelligenza artificiale e discriminazioni di genere. – 2. La discriminazione algoritmica e i suoi diversi fattori. – 3. Il linguaggio degli algoritmi e l'impatto discriminatorio. – 4. Lavoro femminile e IA. – 5. AI Act europeo e diritti delle donne. – 6. Considerazioni sul Trattato del Consiglio d'Europa. – 7. Una nuova sfida: l'IA americana e il “*free speech*” algoritmico. – 8. Postilla: la legge italiana e il riferimento specifico alla “parità di genere”.

*1. Intelligenza artificiale e discriminazioni di genere.*

L'intelligenza artificiale (IA) sta impattando su quasi ogni aspetto della vita delle persone e delle istituzioni, ponendo enormi interrogativi sui diritti fondamentali e sulla necessità di strumenti efficaci per governarne l'impatto. Come ogni rivoluzione tecnologica, quella dell'Intelligenza artificiale contiene enormi potenzialità, ma anche rischi innumerevoli per la garanzia dei diritti fondamentali. Infatti, grazie alla possibilità di analizzare enormi quantità di dati, prendere decisioni autonome e apprendere autonomamente, i sistemi di intelligenza artificiale stanno rapidamente penetrando in settori come la sanità, la mobilità, l'efficienza amministrativa e l'accessibilità dei servizi pubblici. Proprio queste caratteristiche possono comportare violazioni sistematiche o involontarie dei diritti umani, specialmente nel campo della privacy, della non discriminazione, della libertà di espressione e del diritto a un equo processo.

Così si sono già posti interrogativi sulla possibile manipolazione della decisione giudiziale che utilizzi algoritmi predittivi o sulla invasione della privacy creata da sistemi di sorveglianza. La natura spesso complessa dei modelli IA rende difficile identificare chiaramente le responsabilità in caso di violazioni di diritti, facendo emergere la necessità di nuovi approcci regolamentari e

giuridici, incentrati sulla trasparenza e la tracciabilità dei percorsi, in modo da poter mantenere, in qualsiasi momento, il pieno controllo umano.

Negli ultimi anni, la necessità di un approccio etico e rispettoso dei diritti fondamentali da parte degli algoritmi e delle ultime frontiere dell'IA, dai sistemi di LLM all'ultima frontiera dell'IA generativa, è stato al centro delle riflessioni e preoccupazioni delle moderne democrazie occidentali, con un approccio che si è negli ultimi tempi divaricato fra le scelte "regolatorie" europee e le decisioni trumpiane di totale indifferenza ai problemi di rispetto dei diritti umani e della trasparenza dei sistemi, come vedremo.

Nella nostra ricerca, ci siamo concentrati su un aspetto specifico, ma centrale e cioè del modo in cui l'IA possa diventare fortemente discriminatoria soprattutto nei confronti delle donne.

Il punto di partenza sono stati proprio gli studi sugli impatti degli algoritmi sulle persone, che hanno evidenziato la capacità fortemente selettiva e discriminatoria nei confronti delle donne, potenziata anche quando al profilo del genere viene sommato quello razziale.

Partiamo dalla vicenda diventata ormai famosa, nella quale una dottoranda all'MIT Media lab, Joy Buolamwini, scopre e denuncia gli algoritmi di riconoscimento facciale di società come IBM, Microsoft e altre grandi aziende, i quali risultavano molto meno accurati sui volti delle persone di pelle scura e ancora meno sulle donne, in generale, e sulle donne di pelle scura, in particolare, rispetto a quanto non lo fossero per gli uomini, in particolare bianchi. Questo progetto, nato con il nome di *Aspire Mirror*, una volta chiarito che i sistemi discriminavano in quanto allenati su maschi bianchi, cambia nome in "*Gender Shades*" e Buolamwini decide di fondare la *Algorithmic Justice League*, un gruppo che intendeva portare alla luce il fenomeno della discriminazione algoritmica, chiamata "*the Coded Gaze*", lo sguardo dell'algoritmo, espressione che intendeva fare riferimento al "*Male Gaze*", e quindi al sistema di cultura patriarcale e maschilista che influenza non soltanto il modo di rappresentare le donne, ma la stessa rappresentazione della realtà.

Il percorso di Buolamwini è influenzato anche da Cathy O'Neil, una matematica e attivista, che nel suo volume *Weapons of Math Destruction* aveva denunciato come i sistemi algoritmici di modelli matematici per orientare alcune decisioni future non fossero per nulla neutrali, ma, al contrario, incorporando dati storici esattamente come i sistemi di database di riconoscimento facciale, fossero spesso sbilanciati, anche inconsapevolmente, in favore dei gruppi sociali più privilegiati.

Il mondo dell'IA, non sottoposto a regole, riproduce in modo veloce e incontrollato quel "monopolio maschile" che fin dalla storia antica rappresenta

un ostacolo culturale fortissimo alla costruzione di una democrazia paritaria e del pieno rispetto del principio di uguaglianza.

## 2. *La discriminazione algoritmica e i suoi diversi fattori.*

Diversi sono i motivi della discriminazione algoritmica, che ci rendono pienamente consapevoli che gli algoritmi non siano affatto “neutri” e che l’intelligenza artificiale, riproducendo quella “umana”, ne incorpora le profonde differenze esistenti, differenze e disuguaglianze che purtroppo discriminano prima di tutto le donne.

In primo luogo, il “sesso” di chi programma, essendo i programmatori in moltissima parte uomini; in secondo luogo, l’incorporazione di *bias* esistenti nella società, ancora fortemente, anche se spesso a livello inconscio, maschilisti; da ultimo, il modo di funzionamento degli algoritmi, che, raggruppando e selezionando, producono impatti ancora più discriminatori e pervasivi di quelli dei dati che hanno assorbito e che incorporavano pregiudizi.

Il primo aspetto, ben presente ormai anche a livello istituzionale, è costituito dalla prevalenza di programmatori uomini, che riproducono la visione maschile anche nel mondo dell’intelligenza artificiale.

I dati evidenziano come il *gender gap* coinvolga anche i settori più innovativi e all’avanguardia in tema di tecnologia e intelligenza artificiale, in cui attualmente solo il 22% dei professionisti su base mondiale è di sesso femminile contro il restante 78% maschile, mentre alla luce dei dati del *Global gender gap* del 2025, in generale la situazione va migliorando e si attesta a un 68% per gli uomini e a un 32% per le donne.

Va anche evidenziato che le donne rappresentino solo il 18% nei principali convegni scientifici in tema di IA e il 20% dei docenti di IA, come che femminile sia soltanto il 15% e il 10% del personale di ricerca dei colossi informatici come Facebook e Google, principali sviluppatori oggi delle ricerche innovative fuori dalle aule universitarie.

Nelle Università, comunque, alcuni settori scientifici, come in particolare Ingegneria, sono ancora dominati dagli uomini: in Italia, solo il 27% dei laureati ingegneri è donna, un dato che presenta comunque miglioramenti rispetto al 23% del 2020, e anche in Paesi come gli Stati Uniti nell’ultimo decennio si è avuta una flessione del 10% delle lauree STEM da parte delle donne. Si tratta infatti di una palese “segregazione orizzontale”, che si verifica quando in un’area vediamo una prevalenza di uno dei due sessi. Sul problema si ipotizzano cause

diverse, a partire da aspetti psicologici e dall'esistenza di stereotipi che pur conosciuti non sono ancora oggetto di politiche pubbliche efficaci.

Ancor oggi il modello di esperto informatico viene rappresentato come un individuo isolato, con scarse capacità relazionali e pochi contatti umani, provocando un allontanamento delle donne dalla scelta di questo tipo di studi, meno attrattivi per una popolazione femminile più incline a prediligere lavori in team o nei quali l'impatto relazionale ed emotivo sia presente. Su questa rappresentazione pesa ancora lo stereotipo che, a partire dal mondo greco, ritiene le donne meno adatte a calcoli o a dimensioni matematiche rispetto agli uomini.

Tale situazione condiziona e condizionerà in futuro, in negativo, la stessa situazione lavorativa delle donne, alla luce della circostanza che oggi come in futuro il possesso di abilità informatiche e tecnologiche consente non solo di avere maggiori possibilità di accesso, ma soprattutto di ambire a posizioni lavorative migliori e meglio retribuite.

La circostanza che a concepire gli algoritmi siano quasi esclusivamente uomini porta quindi con sé il concreto rischio che essi immagazzinino stereotipi di genere e che tali stereotipi si riproducano e assumano dimensioni più ampie nel momento stesso in cui vengano utilizzati.

D'altronde i dati del *Global Gender Gap Report*, pubblicato ogni anno dal *World Economic Forum* e basato sul confronto fra 149 Paesi in relazione ai progressi sulla parità di genere certificano questa situazione preoccupante in modo inequivocabile. I pregiudizi implementati da algoritmi costruiti da uomini stanno già discriminando le donne, ad esempio, sul reclutamento dei candidati per determinate posizioni di lavoro. Questo è accaduto, ad esempio, nel caso dell'algoritmo che Amazon ha dovuto abbandonare, proprio perché favoriva sistematicamente candidati uomini nella selezione dei *curricula* in quanto era stato "allenato" e programmato tramite l'immissione di *curricula* prevalentemente maschili; lo stesso problema si è verificato anche nel campo del c.d. *platform work*, come nel caso di Uber.

Anche la piattaforma nota come LinkedIn ha dovuto fronteggiare problemi di "discriminazione automatizzata" quando le offerte lavorative per profili altamente retribuiti venivano visualizzate maggiormente nelle ricerche effettuate da utenti di sesso maschile a discapito di quelli di sesso femminile; lo stesso vale per le piattaforme di assistenza virtuale come Siri sono state tacciate di sessismo per l'associazione tra femminilità e servilismo che contribuisce a rafforzare gli stereotipi di genere, come evidenziato dal *World Economic Forum*.

Come abbiamo appena visto, risulta evidente che algoritmi costruiti da uomini, che introducono elementi o dati di per sé discriminatori, si presentino a loro volta discriminatori.

L'intelligenza artificiale presenta, infatti, un potenziale e rilevante problema di diseguaglianza di genere. Gli algoritmi funzionano, infatti, aggregando i dati che l'uomo fornisce loro: se tali dati risultano condizionati da squilibri o pregiudizi è inevitabile che la macchina discrimini determinate categorie.

Sui rischi, in termini di lesione dei diritti fondamentali, derivanti dall'utilizzo di algoritmi predittivi in ambito giudiziario, con particolare riferimento alla materia penale, emblematico fu negli Stati Uniti il c.d. "caso *Loomis*", del 2016, inquadrabile, probabilmente, come *leading case*. In quella circostanza l'imputato era stato infatti condannato a sei anni di reclusione dal Tribunale del Wisconsin sul mero risultato fornito da un algoritmo predittivo denominato *Compas*, concepito per valutare il rischio di recidiva di un soggetto e di cui tuttavia non era dato conoscere il funzionamento. L'imputato, sostenendo che fosse stato leso il proprio diritto ad un equo processo, decise così di ricorrere alla Corte Suprema del Wisconsin, che tuttavia rigettò il ricorso, affermando, pur con alcune cautele, come fosse legittimo l'utilizzo dell'algoritmo menzionato nel processo di autodeterminazione del giudice.

Al contrario, nel caso *K.W. v. Armstrong*, del 2023, la Corte distrettuale dell'Idaho ha sottolineato la necessità di garantire, in corrispondenza al canone – fondamentale – del giusto processo, l'accesso alle informazioni relative alla formula algoritmica che governava l'erogazione dell'assistenza sanitaria per persone con disabilità. Allo stesso modo, nel più risalente caso *Houston Federation of Teachers, Local 2415 v. Houston Independent School District*, del 2017, i giudici hanno affermato che l'utilizzo di algoritmi segreti per l'adozione di decisioni in merito all'assunzione di decisioni gravanti sulla carriera degli insegnanti da parte della *Houston Independent School District* contrastava con i principi del giusto processo. Diverso è il caso *People v. Wakefield*, del 2022, dove è stato negato alla difesa di accedere alle formule algoritmiche utilizzate per l'interpretazione del DNA.

La riproduzione degli stereotipi può avvenire non solo con l'immissione "intenzionale" di informazioni affette da pregiudizi, ma anche, in via indiretta e spesso inconsapevole, tramite informazioni a prima vista "neutre", che, però, in un secondo momento, rielaborate dalla macchina stessa tramite meccanismi di *machine learning* e *deep learning*, producono un effetto discriminatorio "in uscita" sulla base del genere.

Moltissimi studi dimostrano, ad esempio, come gli stereotipi si annidino anche già nel solo nel linguaggio e che ciò ha notevoli implicazioni sul funzionamento degli algoritmi, i quali, avendo associato nel corso dell'apprendimento di una lingua un determinato genere a determinate caratteristiche, tenderanno a perpetuare tale assimilazione al momento della loro utilizzazione pratica, con effetti potenzialmente discriminatori.

Le ragioni che determinano il funzionamento discriminatorio delle tecniche di intelligenza artificiale sono da ricondurre al ruolo assolto dai dati.

L'intelligenza artificiale discrimina, infatti, perché si fonda e si alimenta di dati, talvolta non completi, volontariamente oppure involontariamente parziali, errati; dati che omettono di riflettere la realtà sulla quale la tecnologia è destinata a riverberare i suoi effetti.

Le conseguenze deteriori dell'intelligenza artificiale si sono dimostrate, però, particolarmente evidenti sul versante specifico della violazione del principio di eguaglianza di genere. Questo aspetto è emerso in modo chiarissimo di fronte all'analisi del funzionamento dei sistemi di riconoscimento facciale, incapaci di riconoscere volti di donna, ma particolarmente accurati rispetto a volti di uomini, specie se bianchi. I dati, in questo caso, si rivelano parziali perché sono forniti alle tecnologie di intelligenza artificiale da programmatori uomini, che omettono di assicurare la costruzione di un *data-set* completo con effetti deteriori sul funzionamento, discriminatorio, degli algoritmi.

Non si tratta, però, soltanto di una discriminazione che deriva dal funzionamento delle tecnologie di intelligenza artificiale perché i dati sono, come detto, non adeguatamente rappresentativi oppure affetti da vizi e parziali. In altre parole, la discriminazione non nasce solo nel momento in cui i programmatori individuano e, poi, selezionano i dati da fornire ai sistemi di intelligenza artificiale. A questa origine o causa, se ne affiancano altre che ripetono le fasi in cui si costruisce l'algoritmo e quelle, particolarmente complesse, in cui se ne articola il funzionamento successivo. In tutti questi "momenti" o passaggi, le tecnologie di intelligenza artificiale hanno dimostrato una attitudine più o meno elevata a produrre effetti discriminatori, proprio perché tendenti a selezionare e a raggruppare, e questa operazione meccanica, se compiuta su un gruppo di dati selezionato in origine in modo differente e quindi spesso discriminatorio, conduce la macchina a moltiplicarne a dismisura gli effetti.

Va anche sottolineato che la ricerca del "dove" si annida il rischio dell'effetto discriminatorio, del "quando" esso può verificarsi e del "come" quest'ultimo si realizza, si rivela particolarmente complessa per tecnologie di intelligenza artificiale di largo impiego e il cui funzionamento diviene autonomo sino ad essere, talvolta, addirittura sconosciuto al programmatore iniziale. In questi casi, la tecnologia agisce senza che chi ne abbia costruito inizialmente l'algoritmo sia in grado di prevedere gli effetti concreti derivanti dal suo funzionamento su larga scala. Ci si riferisce alla c.d. *black box theory* che evoca l'idea per la quale, una volta programmato, istruito, messo in funzione, l'algoritmo sarebbe capace di agire secondo modalità che divengono progressivamente ignote al programmatore. L'"apprendimento" che pure la persona impartisce alla tecnica per

assicurane un funzionamento corrispondente alle finalità assumerebbe, in altre parole, una connotazione progressivamente distante ed imprevedibile, rendendo oscuri i passaggi che consento a quella tecnica di raggiungere un risultato specifico. Detto altrimenti, il funzionamento delle tecnologie di intelligenza artificiale non sarebbe conoscibile e ricostruibile *ex post*, nemmeno da chi le abbia programmate, essendo viceversa ignoto come questi sistemi producono gli effetti originariamente voluti e sempre che questi ultimi corrispondano poi effettivamente a ciò che era stato originariamente desiderato.

Alcuni sistemi di intelligenza artificiale sono, inoltre, “chiusi” rispetto ad interventi successivi esterni, cioè ad azioni del programmatore, precludendo correzioni al malfunzionamento delle tecnologie ed ostacolando comportamenti distorsivi e discriminatori. Ed è per queste ragioni che l’aggiunta di dati ad un *data-set* iniziale oppure la previsione del controllo umano, il c.d. *human in the loop*, non è sempre sufficiente a correggere il funzionamento del sistema, escludendo discriminazioni derivanti dall’intelligenza artificiale. Ciò, soprattutto, come detto, perché quando non si conosce esattamente come il sistema raggiunge l’obiettivo desiderato, diventerà praticamente impossibile adottare comportamenti che rimedino al malfunzionamento del sistema. Tanto più, quindi, l’intelligenza artificiale agisce in modo autonomo rispetto alla programmazione della persona e tanto più le fasi entro cui si snoda il suo funzionamento è oscuro, tanto più l’individuazione del rischio che si producano effetti discriminatori diviene importante, di complessa prevenzione, di difficile contrasto.

### 3. Il linguaggio degli algoritmi e l’impatto discriminatorio.

Come è noto, il problema del linguaggio è molto importante per il suo impatto sui diritti. Il linguaggio non solo esprime la realtà, ma la può anche trasformare e condizionare, per la sua natura performativa. In Italia, negli anni recenti, a livello istituzionale si è molto discusso, soprattutto per l’iniziativa del dizionario Treccani, che ha adottato un linguaggio c.d. binario, rigettando l’universale maschile come scelta necessaria per la lingua italiana. La natura performativa del linguaggio, poi, può incidere molto negativamente sulla realtà, soprattutto per ciò che riguarda il c.d. hate speech, specie se utilizzato con strumenti informatici, in quanto potenzialmente più lesivi a causa della loro “virilità”.

Rispetto a tali tematiche, il campo di indagine rispetto alle tecnologie di IA più interessante è senza dubbio quello dei sistemi di c.d. *natural language processing* (NLP).

Il *natural language processing* si occupa di studiare e di costruire meccanismi di intelligenza artificiale in grado di permettere che i computers comprendano testi e parole, replicando abilità umane.

Si tratta, quindi, di algoritmi che si vorrebbero in grado di analizzare, rappresentare e comprendere il linguaggio naturale “delle persone”, sia esso parlato oppure scritto. Come segnalato anche nel Report UNESCO del 2024, dal titolo *Challenging systematic prejudices: an investigation into bias against women and girls in large language models* emerge, allora, in questa prospettiva, il problema dell'effetto che i pregiudizi e gli stereotipi che caratterizzano il linguaggio, scritto e parlato, riflettono sui meccanismi di intelligenza artificiale, tra cui non fanno eccezione i citati sistemi di *natural language processing*.

Un esempio delle ripercussioni discriminatorie che caratterizzano questi peculiari sistemi di intelligenza artificiale è costituito dagli strumenti di traduzione automatica.

Cercando di adattare un testo redatto in una lingua neutra sotto il profilo del genere (si pensi al caso dell'inglese), ad una lingua connotata, viceversa, dalla esistenza della doppia declinazione maschile/femminile (l'italiano oppure il francese e, ancora più complesso, il tedesco per la presenza del neutro), tali sistemi assai di frequente tendono ad oscurare la dimensione di genere per favorire l'impiego del già citato c.d. maschile universale, altrimenti definito “maschile neutro”. Il “maschile neutro”, cioè, viene considerato idoneo a riflettere le differenze di genere, che sono ritenute superflue da riprodurre in sede di traduzione del testo. Si tratta di una semplificazione linguistica che, però, come già detto, nega la rappresentazione della realtà, fatta di uomini e di donne.

Il fenomeno, peraltro, riproduce anche quanto è in tutto identico avviene ancor oggi, almeno in Italia, nel linguaggio comune dove nella maggior parte dei casi si ricorre al maschile poiché ritenuto presuntivamente idoneo a “coprire” anche il femminile.

Nel simulare il linguaggio comune o non mediato artificialmente, gli strumenti di *natural language processing*, come sono appunto i *softwares* di traduzione automatica, riflettono così i *bias* e, sarebbe meglio dire, gli “errori” del linguaggio ordinario.

I sistemi di *natural language processing* ripropongono gli usi della lingua, discriminatori o sessisti di cui si è detto e, tuttavia, possono anche essere utilmente impiegati anche per finalità di tipo diverso. Si tratta, cioè, di un diverso e opposto utilizzo di tali sistemi allo scopo di promuovere e diffondere un linguaggio che si dimostri, viceversa, sensibile alle differenze di genere. Sul punto, un esempio significativo è offerto da Google, che, nel 2017, ha scelto di ridefinire il proprio programma di traduzione automatica allo scopo di inserire una

gamma di traduzioni alternative capaci di rispondere alle specificità di genere.

L'esempio di Google Translate è molto interessante perché dimostra che, laddove sussista una volontà umana in questo senso, l'intelligenza artificiale è in grado di passare dall'essere un mezzo per discriminare al rappresentare uno strumento, viceversa, capace di contrastare la diffusione e l'irrobustimento di stereotipi. In questo senso, un utilizzo consapevole e trasparente dell'intelligenza artificiale, rispettoso dei principi costituzionali, potrebbe permettere di invertire la rotta favorendo un impiego della innovazione tecnologica che non sia "nemica" dei diritti delle persone, ma sappia invece contribuire alla loro tutela.

La doppia natura dell'intelligenza artificiale – da un lato, mezzo (in)consapevole di discriminazione e, dall'altro, potenziale alleato per assicurarne il contrasto – traspare in modo evidente se si considera il fenomeno del c.d. odio online.

Se l'intelligenza artificiale assolve ad un ruolo importante nella creazione e nella diffusione dell'odio online, al tempo stesso essa può, però, atteggiarsi anche a strumento efficace per il suo contrasto agendo in modo analogo a quanto già osservato trattando dei sistemi di traduzione automatica.

Se parliamo di odio online, d'altra parte, occorre sempre considerare che le prime vittime sono proprio le donne.

È, infatti, innegabile che il web, lasciato nella piena disponibilità degli utenti e sprovvisto di una regolamentazione, costituisca un terreno fertile per la genesi di comportamenti discriminatori, che assumono la forma dei c.d. discorsi d'odio. La circostanza che il web sia luogo facilmente dominabile dai c.d. "odiatori" (*haters*) non coincide con l'esistenza di una responsabilità dell'intelligenza artificiale quale creatrice e promotrice dell'odio online, se non nella misura in cui essa si sottragga all'apprestamento di argini sufficienti alla diffusione di contenuti discriminatori.

È noto che anche l'intelligenza artificiale contribuisce fortemente al fenomeno dell'odio online, a causa della c.d. indicizzazione dei contenuti impiegati sul web: le piattaforme online e, in particolare, i social networks funzionano, infatti, sulla base del meccanismo della c.d. profilazione attraverso cui vengono monitorati e analizzati i comportamenti degli utenti al fine di prevederne comportamenti, preferenze e scelte, sì da fornire contenuti personalizzati e, nelle intenzioni, corrispondenti ai desideri, ai gusti di coloro che fruiscono delle piattaforme.

Si è infatti dimostrato che contenuti viziati da sentimenti d'odio sono messi a disposizione in modo proporzionalmente maggiore in favore di chi si è già dimostrato ad essere più incline a ricorrere a contenuti discriminatori, favorendo la creazione di quello che la dottrina chiama la c.d. "camera dell'eco",

cioè l'amplificazione e il rafforzamento dell'opinione che si forma all'interno di un gruppo con il successivo orientamento dello stesso verso posizioni progressivamente più estreme e radicalizzate.

Secondo questo meccanismo, un contenuto discriminatorio oppure un fenomeno d'odio può essere amplificato da algoritmi, volutamente programmati per assicurare che i contenuti mostrati all'utente siano in linea con le sue preferenze.

A questo, si affianca un secondo aspetto meritevole di considerazione.

Come già nella realtà, più un contenuto è "d'impatto" e dirompente, tanto più tende a suscitare interesse. Ne discende che, in tanto gli algoritmi sono programmati per dare massima diffusione a ciò che presenta più interesse, nel momento in cui il contenuto che suscita più interazioni tra gli utenti è espressione di odio e di intolleranza, se la piattaforma non vi pone argini, sarà questo stesso contenuto ad essere diffuso fra gli utenti grazie all'azione dei sistemi di intelligenza artificiale

Si spiegano, così, fenomeni purtroppo all'ordine del giorno come il c.d. *revenge porn*, lo *stalking* digitale e, più in generale, il sessismo e la discriminazione che vengono amplificati dalla rete e destinati a diventare virali in poche ore. La "viralità" è, del resto, caratteristica intrinseca alla rete e obiettivo a cui contestualmente ambiscono le piattaforme. Quando però virale è un contenuto negativo e di odio, occorre immaginare antidoti efficaci alla diffusione di tali contenuti: per questo è importante ragionare sulla regolamentazione e sulla moderazione dei contenuti disponibili sul web, una condizione necessaria perché la rete e, dunque, l'intelligenza artificiale che la governa, passi dall'essere solo creatrice di odio oppure, anche, un'alleata efficace per il suo contrasto.

Come osservato, se non si interviene su algoritmi "viziati", questi agiranno in modo automatico e ne diffonderanno i contenuti, anche, quelli negativi: le tecnologie di intelligenza artificiale non sono però soltanto al servizio dell'odio online, dal momento che esistono e possono essere agilmente sviluppate modalità e strategie che consentono all'intelligenza artificiale di arginare la diffusione dell'odio sul web.

L'intelligenza artificiale può, così, diventare un valido alleato nella lotta alla diffusione dell'*hate speech online* soprattutto mediante l'utilizzo di strumenti di moderazione automatizzata, che si stanno progressivamente sostituendo alle attività prima assolute in via esclusiva dai moderatori umani.

Attualmente, le piattaforme online dispongono di una pluralità di strumenti algoritmici attraverso cui filtrare le c.d. *bad information* diffuse online, siano esse costituite da immagini, video oppure testi. Il filtraggio dei contenuti negativi o discriminatori avviene attraverso le già più volte citate tecniche di *machine*

*learning*, ossia mediante la programmazione di algoritmi capaci di riconoscere i contenuti lesivi per poi, successivamente, limitarli oppure, nei casi più gravi, eliminarli.

Nonostante la necessità di utilizzare sistemi automatizzati per arginare efficacemente la diffusione di odio on line, occorre considerare che tale manipolazione contiene un rischio, molto rilevante per i valori delle democrazie occidentali, che si reggono sulla libertà di manifestazione del pensiero e del rifiuto di qualsiasi forma di censura: il rischio infatti è quello di colpire in modo indiscriminato vere espressioni o pensieri di odio, insieme invece a espressioni o contenuti che magari sono soltanto di critica, limitando quindi e standardizzando il pensiero delle persone e il dibattito pubblico.

Gli strumenti di moderazione automatica rischiano, quindi, da una parte, di essere troppo permissivi e di non raggiungere l'obiettivo desiderato, cioè la tutela dei diritti degli utenti; dall'altra, di essere troppo stringenti, limitando il diritto alla libertà di espressione anche di chi non commette alcun tipo di abuso.

Una garanzia in questa direzione può certamente derivare dalla possibilità di conoscere a priori i criteri sulla base dei quali l'intelligenza artificiale opera quando monitora i contenuti diffusi sulle piattaforme online a cui dovrebbe affiancarsi la capacità di conservare il controllo sul funzionamento successivo dell'algoritmo. Ad oggi, i sistemi di moderazione, così come più in generale i sistemi di intelligenza artificiale, non sempre rispondono al principio di trasparenza, anche poiché gli algoritmi sono spesso coperti da garanzie di proprietà intellettuale. La non trasparenza del funzionamento dei sistemi di intelligenza artificiale ne rende, così, particolarmente arduo un controllo esterno impedendo agli utenti di conoscere in ultima istanza come opera la piattaforma anche per quanto attiene alla moderazione dei suoi contenuti.

In considerazione della difficoltà di operare un controllo diretto sugli algoritmi, una seconda strategia di intervento in grado di bilanciare diritti e libertà potrebbe essere costituita da un intervento *ex ante*, individuando a monte i principi a cui i programmatori e le programmatrici devono obbligatoriamente conformare le rispettive mansioni.

Proprio per rispondere a queste problematiche, rese evidenti da una serie di attentati terroristici alimentati anche dall'odio sul web, nel 2016 la Commissione europea, insieme con alcune delle più influenti aziende informatiche, ha adottato il Codice di condotta per lottare contro le forme illegali di incitamento all'odio online.

Nel Codice di condotta vengono responsabilizzate le piattaforme online sia con riguardo all'eliminazione di contenuti lesivi dei diritti fondamentali, sia quanto alla opportunità di istruire gli utenti in ordine alle conseguenze legate

alla diffusione di contenuti discriminatori e potenzialmente pericolosi. Il Codice di condotta mira così ad agire su un duplice fronte: la rimozione coattiva dei contenuti e l'“educazione” degli utenti mediante il coinvolgimento attivo dei *providers* digitali.

In questa prospettiva, è anche determinante il ruolo degli utenti, che possono a loro volta segnalare contenuti potenzialmente discriminatori, attivando strumenti di moderazione delle piattaforme.

Questa modalità di contrasto dell'odio online, poggiando sul ruolo assolto dalle persone che si interfacciano con le piattaforme digitali, non fa che dimostrare ancora una volta come la via per il contrasto dell'odio online non possa che passare da una alleanza fra la persona e i sistemi di intelligenza artificiale che si riveli attenta garante dei diritti costituzionali degli utenti del web.

Il 25 gennaio 2025 il codice di condotta riveduto è stato integrato nel quadro normativo della legge sui servizi digitali e ad esso hanno aderito numerose aziende informatiche, come Facebook, Microsoft, Youtube, Instagram.

#### 4. *Lavoro femminile e IA.*

Con l'IA il mercato del lavoro si sta trasformando profondamente, consentendo da una parte modalità organizzative più agili e positive – come lo smart working – per chi, come soprattutto le donne, si trova a conciliare i tempi del lavoro con i tempi della cura familiare, dall'altra, introducendo nuove figure di lavoratori, con competenze tecniche molto richieste e potenzialmente avvantaggiate.

Questa trasformazione, in ogni caso, presenta anche molti rischi, soprattutto per le donne che sono a livello mondiale meno occupate, più dedite a lavori esecutivi e meno competenti, come abbiamo detto, a livello tecnologico e informatico.

Alcuni impieghi che richiedevano mansioni esecutive sono stati depotenziati o eliminati, colpendo in modo rilevante le categorie di persone in età più avanzata, difficili da istruire sulle nuove competenze digitali richieste.

Ancora una volta, l'impatto sulle donne è potenzialmente molto più forte: secondo un Rapporto del 2025 dell'Organizzazione Internazionale del Lavoro, nei Paesi ad alto reddito l'IA mette a rischio il 9,6 % dei posti di lavoro occupati da donne, rispetto al 3,5% di quelli occupati da uomini. Il dato è impressionante e non fa altro che confermare quanto gli esperti avevano già previsto, alla luce delle caratteristiche del lavoro femminile, meno tecnologico e, in generale, più precario.

La trasformazione digitale costituisce dunque una profonda sfida per la nostra società e, soprattutto, per le donne, e il pericolo che nelle grandi trasformazioni epocali queste ultime 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.

Di questo, sul piano mondiale ci si sta rendendo conto e, per fortuna, da anni ci sono state proposte sul piano generale, come quelle evidenziate in un documento risalente, ma ancora molto attuale: un report intitolato *Artificial intelligence: open questions about gender inclusion*, Policy Brief – W20 Argentina 2018, nel quale si incoraggiavano i governi di tutti gli Stati a coinvolgere le donne nella costruzione degli algoritmi, con programmi specifici di formazione, adottando linee guida che ne favorissero la trasparenza e la tracciabilità, con attenzione anche agli impatti presenti e futuri della trasformazione digitale sul mondo del lavoro.

Di tutte queste problematiche è stata consapevole soprattutto l'Unione Europea, nella quale fin dal 2018, è stata adottata una strategia per lo sviluppo di un'intelligenza artificiale antropocentrica, seguita da un piano coordinato in materia, da linee guida sull'etica nella materia, fino all'adozione di un vero regolamento basato sul criterio del rischio algoritmico e sulla verifica dell'impatto sui diritti fondamentali nell'ambito dei settori più rilevanti nella vita delle persone, come quello della sicurezza, della giustizia, della sanità.

### *5. AI Act europeo e diritti delle donne.*

Con un percorso complesso, l'Unione Europea ha creato un quadro normativo chiaro e uniforme per l'utilizzo dell'intelligenza artificiale, ispirato a principi per garantire la sicurezza, la trasparenza e il rispetto dei diritti fondamentali. Il "Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i Regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le Direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828", noto come Regolamento sull'intelligenza artificiale o "Artificial Intelligence Act", è entrato in vigore il 1° agosto 2024, stabilendo una graduale entrata in vigore dei differenti articoli: l'art. 5 sui sistemi vietati è entrato in vigore il 1 febbraio del 2025, mentre le disposizioni relative ai sistemi a rischio elevato dovrebbero diventare pienamente applicabili dal 2 agosto 2026, al termine del periodo di adeguamento biennale.

Un aspetto molto discusso dal Regolamento è stato proprio quello dell'impatto che i sistemi di IA possono avere sui diritti di genere, proprio per evitare quei *bias* e quelle discriminazioni che colpiscono soprattutto le donne.

Nel Preambolo e nel testo dell'art. 27 sui sistemi "ad alto rischio" si fa riferimento alla *compliance* rispetto ai diritti fondamentali, quali contenuti anche nelle fonti europee e, in particolare, nei Trattati e nella Carta dei diritti fondamentali: la parità di genere costituisce un pilastro dell'Unione, tanto da essere oggetto non solo di specifici regolamenti, direttive e atti, ma anche di una intensa giurisprudenza della Corte di Giustizia, che vede una dimensione molto significativa ed estesa del principio di "non discriminazione di genere", applicato in ambiti vasti e non limitati al diritto del lavoro.

Oltre a questo aspetto, va sottolineato che lo stesso art. 5 sui sistemi vietati costituisce una formulazione generale di aspetti discriminatori soprattutto nei confronti delle donne: ci riferiamo alla parte che riguarda il divieto dei sistemi di profilazione, che, come è noto, erano stati utilizzati soprattutto per la selezione del personale e hanno costituito casi di evidente discriminazione di genere.

## 6. *Considerazioni sul Trattato del Consiglio d'Europa.*

Nel panorama internazionale troviamo anche un altro importante fondamentale documento regolatorio: si tratta del primo Trattato al mondo sul tema, approvato dal Consiglio d'Europa, frutto del lavoro di oltre due anni di un comitato intergovernativo che ha coinvolto attivamente i 46 Stati membri del Consiglio d'Europa, l'Unione Europea e altre 11 nazioni non appartenenti all'organizzazione (Argentina, Australia, Canada, Costa Rica, Santa Sede, Israele, Giappone, Messico, Perù, Stati Uniti d'America e Uruguay).

Il Trattato, aperto anche alla firma di nazioni non europee, delinea un quadro normativo completo che abbraccia l'intero ciclo di vita dei sistemi di IA. Non solo ne definisce i criteri di progettazione, sviluppo e dismissione, ma affronta anche i potenziali rischi connessi, promuovendo al contempo un'innovazione responsabile.

Emerge, fin dal Preambolo, una chiara consapevolezza delle opportunità offerte dai sistemi di intelligenza artificiale, per lo sviluppo del benessere dell'umanità, e anche dei diritti fondamentali, ma al tempo stesso dei rischi che essi possono arrecare alla stessa dignità umana, ai diritti, alla democrazia e allo stesso Stato di diritto.

Nel Preambolo viene evidenziata la necessità di evitare che gli stessi sistemi

di intelligenza artificiale possano essere utilizzati a fini repressivi, attraverso arbitrarie e illegittime pratiche di sorveglianza o di censura, lesive del diritto di privacy individuale.

Il Trattato individua un quadro normativo completo che abbraccia l'intero ciclo di vita dei sistemi di IA: non solo ne definisce i criteri di progettazione, sviluppo e dismissione, ma affronta anche i potenziali rischi connessi, promuovendo al contempo un'innovazione responsabile.

Da una prima analisi, sembra che la Convenzione adotti un approccio analogo all'AI Act in termini di regolamentazione delle tecnologie di IA in base al rischio e con riferimento all'opportunità di identificare in chiave preventiva i rischi associati alle tecnologie di IA e, successivamente, di identificare strategie di mitigazione e contenimento dei danni derivanti dal ricorso all'IA.

Rispetto però al regolamento europeo, la Convenzione insiste nella necessità di una piena padronanza dei sistemi di intelligenza artificiale allo scopo (art. 3) di garantire il rispetto dei «diritti umani, democrazia e Stato di diritto». Da ciò, l'ampiezza del testo, che si rivolge sia alle autorità pubbliche, che ai soggetti privati che li utilizzano.

Un intero capitolo (cap. III) è dedicato ai principi che dovrebbero governare il ricorso alle tecnologie di IA: l'impostazione è sicuramente simile a quella ricavabile dall'AI Act, ma, come si è già detto, si riscontra una più esplicita attenzione ai diritti umani soggetti ad essere negativamente pregiudicati dalle tecniche di IA.

Fra i principi fondamentali della Convenzione troviamo, come nell'AI Act europeo, la trasparenza e la supervisione. Il testo definisce requisiti specifici in base ai contesti e ai rischi potenziali, come la necessità di una identificazione dei contenuti generati dai sistemi di IA, aspetto importante per la lotta ai *deep fake*.

Inoltre, il Trattato prevede l'obbligo per gli Stati ratificanti di adottare misure per identificare e valutare i rischi (chiedendo un'analisi preventiva per individuare i potenziali pericoli associati all'utilizzo dei sistemi di IA); prevenire e mitigare i rischi (esigendo l'implementazione di strategie per ridurre al minimo le conseguenze negative); garantire responsabilità e obblighi di rendicontazione (individuando i responsabili in caso di impatti negativi e l'adozione di misure per prevenirli); promuovere l'equità e la non discriminazione (dovendo essere assicurato da parte dei sistemi di IA il rispetto dei principi di uguaglianza, inclusa la parità di genere, e il divieto di discriminazione); proteggere la privacy (un aspetto che ricorre e sempre collegato anche al rischio di abuso da parte dei poteri pubblici).

Va infine sottolineato che in questo documento, ancor più che nell'AI Act

dell'Unione, è contenuto un riferimento esplicito alla parità di genere, e non soltanto al principio di uguaglianza.

### 7. *Una nuova sfida: l'IA americana e il "free speech" algoritmico.*

Il 23 luglio 2025, la Casa Bianca ha pubblicato un nuovo documento strategico intitolato *America's AI Action Plan*. Si tratta di un testo lungo, articolato e denso di proposte operative il quale, prima ancora che per i contenuti, colpisce per il linguaggio. Sin dalle prime righe si capisce che non si tratta di una semplice *roadmap* tecnologica ma di una dichiarazione di principio, frutto di una precisa scelta politica ideologica e identitaria.

Il Governo americano dichiara, infatti, che l'intelligenza artificiale è lo strumento attraverso cui gli Stati Uniti devono riaffermare il proprio «dominio tecnologico incondizionato» sul piano globale. Il lessico è muscolare, competitivo, quasi bellico, si parla di corsa, supremazia, infrastrutture da costruire a ogni costo e in tutto il discorso spicca un'espressione ricorrente: "American values".

Opposta la visione rispetto a quella adottata fino a quel momento negli Stati Uniti, dove il centro della riflessione era occupato, come in Europa, da principi universali come i diritti umani, la non discriminazione, la protezione dei soggetti vulnerabili: ora il baricentro si sposta sui "valori americani" e sulla "libertà di parola", da intendere però come opposizione a ogni forma di moderazione, filtraggio o ingegneria sociale. In modo esplicito, il piano chiede di rimuovere dalle linee guida federali ogni riferimento a disinformazione, inclusione, cambiamento climatico.

Come per altri Decreti presidenziali aventi ad oggetto vere e proprie inversioni di tendenza rispetto a temi come il genere, la *diversity*, il *climate change*, la strategia sull'IA utilizza la leva economica.

Il piano prevede, infatti, che gli stati federati che adottano regolazioni sull'IA considerate troppo "onerose" possano essere esclusi dai finanziamenti pubblici. Una misura che suona come un chiaro avvertimento: chi tutela troppo, resta indietro. A essere penalizzati non sarebbero solo gli stati più progressisti, ma anche le aziende e le università che operano in ambienti regolati.

Ancora più delicato è il passaggio dedicato alla protezione delle innovazioni commerciali e governative, che lascia intravedere un potenziale ampliamento del controllo statale sulla diffusione di tecnologie AI, con logiche protezionistiche e di segretezza difficilmente compatibili con i principi di apertura, interoperabilità e trasparenza promossi fino a quel momento sul piano globale all'interno del mondo occidentale.

Sotto il profilo del lavoro, l'*AI Action Plan* parla esplicitamente di «potenziare la *workforce* americana», concentrandosi quasi esclusivamente su aspetti tecnici, come il riferimento a persone formate da impiegare nell'espansione della manifattura e nella costruzione di *data center* e impianti energetici. Nulla si dice, invece, sull'impatto dell'automazione cognitiva nei servizi, nell'amministrazione, nei settori ad alto impiego impiegatizio. Il rischio di perdita massiccia di posti di lavoro viene in questo caso esplicitamente ignorato.

Infine, un capitolo centrale riguarda il ruolo del Dipartimento della Difesa, che viene spinto ad adottare su larga scala tecnologie AI per scopi operativi e decisionali, in un'ottica di automazione crescente dei processi militari e di intelligence. Anche qui, nessun riferimento a vincoli di diritto internazionale, a salvaguardie etiche, a limiti d'uso.

Ma l'aspetto che più interessa nella nostra prospettiva è, ancora una volta, il riflesso sui diritti delle donne: nel documento esplicitamente si vieta, come già era stato fatto nei decreti presidenziali di inizio mandato, di utilizzare prospettive legate alla *diversity, women's rights, human rights, climate change*. E si minaccia anche qualsiasi ente o istituzione, a partire dalle Università, di fare adozione di questa prospettiva e impostazione. Ciò significa che, ad esempio, negli Stati Uniti dovrebbero essere rimossi (alcune piattaforme lo stanno già facendo) quei meccanismi di filtro dell'odio online o dell'adozione indiscriminata di algoritmi contenenti *bias*, come abbiamo detto pregiudizievole soprattutto per le donne.

Emerge allora in modo chiarissimo che fra Stati Uniti ed Europa si sta alzando una vera e propria barriera, che attiene al modo di valutare i diritti fondamentali delle persone: ma è significativo che, per le prospettive più avanzate, come quella del mondo *onlife*, la battaglia veda al centro proprio e ancora i diritti delle donne.

#### 8. *Postilla: la legge italiana e il riferimento specifico alla "parità di genere".*

Nelle more della pubblicazione di questo lavoro è entrata in vigore la legge italiana n. 132/2025 ("Disposizioni e deleghe al Governo in materia di intelligenza artificiale"), un elaborato ampio e complesso composto di 28 articoli che toccano non soltanto i principi e il metodo di utilizzo, ma una serie di materie e ambiti strategici, dalla sanità, al lavoro, alla *cybersecurity*, nonché aspetto essenziali di *governance*.

Per dare un giudizio complessivo occorrerà attendere i decreti attuativi, ai quali, nello stesso testo, è delegato il Governo.

Sulla carta, la disciplina si presenta di grande impatto: essa contiene innanzitutto, come abbiamo detto, una serie di principi e di regole nell'ambito della cornice europea, a cui la legge fa puntuale riferimento, a partire dall'art. 2 (che si occupa di definire i sistemi di IA) e dall'art. 3 (che introduce i principi generali).

Va osservato che quest'ultimo è chiarissimo nella piena adesione all'*AI Act* europeo, richiamandosi alla necessità del rispetto dei diritti fondamentali, valorizzando la Costituzione italiana e il diritto dell'Unione Europea. Sempre sul solco delle norme europee, si ribadiscono come centrali i principi di «trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, accuratezza, non discriminazione, parità dei sessi e sostenibilità» (art. 3, comma 1).

Viene anche ribadito, ai commi 2 e 3 della stessa disposizione, la necessità di trasparenza, correttezza, qualità e appropriatezza dei processi di sviluppo per finalità generali, sottolineando la necessità del «rispetto dell'autonomia e del potere decisionale dell'uomo» e dell'assicurazione di possibilità della «sorveglianza e dell'intervento umano».

Rispetto all'*AI Act*, però, il comma 4 dell'art. 3 si sofferma sulle implicazioni e sui pericoli dell'uso dell'intelligenza artificiale a livello democratico, richiedendo che venga garantito «lo svolgimento con metodo democratico della vita istituzionale e politica (...)» e che non venga pregiudicato «la libertà del dibattito democratico da interferenze illecite, da chiunque provocate, tutelando gli interessi della sovranità dello Stato (...)».

Inoltre, il testo contiene un riferimento esplicito anche alla parità di genere.

Un altro aspetto molto significativo e originale rispetto alle regole europee, che pure ad esso fanno riferimento, è il comma 7 dell'art. 3, che introduce la necessità di garantire alle «persone con disabilità il pieno accesso ai sistemi di intelligenza artificiale e alle relative funzionalità ed estensioni», alla luce del principio di eguaglianza e di non discriminazione.

Molto interessante risulta l'art. 5 della legge (principi in materia di sviluppo economico) che affida allo stato e alle altre autorità pubbliche il compito di sviluppare il mondo produttivo e l'innovazione attraverso la promozione e l'utilizzo dell'intelligenza artificiale come «strumento per migliorare l'interazione uomo-macchina nei settori produttivi» e per «avviare nuove attività economiche» (lett. a), nonché a favorire la «creazione di un mercato dell'intelligenza artificiale innovativo» (lett. b), dove lo Stato dovrebbe facilitare anche la comunità scientifica (lett. c), privilegiare la creazione e utilizzo di data center nazionali (lett. d), nonché favorire la «ricerca collaborativa tra imprese, organismi di ricerca e centri di trasferimento tecnologico in materia di intelligenza artificiale» (lett. e).

L'insistenza sull'autonomia nazionale dei *data center*, in un quadro completamente diverso dal punto di viste della relazione con i principi fondamentali che tutelano le persone, richiama però la strategia americana, fondata proprio sulla difesa del proprio mondo virtuale.

Molte sono le norme dedicate al mondo della sanità e agli sviluppi della robotica (artt. 7-10), con riferimenti alla disabilità (art. 7), alla ricerca e sperimentazione in ambito scientifico (per finalità di prevenzione, diagnosi e cura di malattie, sviluppo di farmaci, terapie e tecnologie riabilitative, realizzazione di apparati medicali, salute pubblica (art. 8), al trattamento dei dati personali (art. 9) e alle implicazioni sullo stesso fascicolo sanitario (art. 10).

Importanti le disposizioni in materia di lavoro (artt. 11-12), di professioni intellettuali (art. 13), della pubblica amministrazione (art. 14) e della giustizia (art. 15).

Nel suo approccio all'intelligenza artificiale, in un mondo diviso sulla strategia europea "di prevenzione e difesa", a cui si contrappone quella americana, che si presenta come una strategia "di guerra", l'Italia propone una terza visione: quello dell'intelligenza artificiale come strategia "di sviluppo". Rispetto agli altri testi esistenti, quello italiano si caratterizza per una consapevolezza anche degli ambiti che verranno maggiormente trasformati dall'utilizzo dell'IA, con un approccio significativo, che oscilla fra riconoscimento delle garanzie e valorizzazione degli ambiti in una chiave di sviluppo.