

**Future of
Workers**

EDIZIONE I

**Le sfide dell'intelligenza
artificiale per una società
più inclusiva**

EDIZIONE 1

TEMA:

Valorizzare diversità ed inclusione

ARGOMENTO:

Le sfide dell'intelligenza artificiale per una società più inclusiva

DIRETTRICE RESPONSABILE:

Maria Laura Fornaci

REDATTORE CAPO:

Tommaso Limonta

COMITATO EDITORIALE:

Claudia Bruno, Bianca Iafelice, Barbara Leda Kenny, Martina Marzi

EDITORE:

Fondazione Giacomo Brodolini

PROGETTO GRAFICO:

Daniela Palumbo

MARZO 2024

Future of Workers

Future of Workers - osservatorio per il lavoro sostenibile è un progetto cofinanziato dal Ministero dell'Ambiente e della Sicurezza Energetica nell'ambito del processo di attuazione della Strategia Nazionale per lo Sviluppo Sostenibile.

FUTURE OF WORKERS

Come sta cambiando il mondo del lavoro con e dopo la pandemia? Come cambierà con la transizione ecologica e digitale?

Future of Workers è l'osservatorio della **Fondazione Giacomo Brodolini** sulle tendenze e le politiche che stanno cambiando i luoghi, le competenze, l'organizzazione del lavoro e quindi la vita delle persone.

La **ricerca** è alla base delle attività della Fondazione Giacomo Brodolini. Crediamo che il cambiamento possa avvenire solo avendo una profonda comprensione delle cose, la passione per realizzarle e la visione politica necessaria a migliorare la vita delle persone.

Vogliamo creare, dunque, consapevolezza, conoscenza e dialogo sui cambiamenti del mercato del lavoro affinché gli obiettivi di sostenibilità della **Strategia Nazionale per lo Sviluppo Sostenibile (SNSvS)** diventino reali opportunità di sviluppo per le istituzioni, le imprese e le persone che lavorano.

L'osservatorio è **tascabile** perché ogni mese offre a professionisti e professioniste, imprese e PA contenuti di facile fruizione per comprendere gli scenari europei in evoluzione.

È **innovativo** e **4.0** perché fornisce strumenti digitali pratici per affrontare insieme a un team di esperti ed esperte le nuove sfide del mercato del lavoro.

Perché un osservatorio sui cambiamenti del mondo del lavoro e sul futuro dei lavoratori?

Perché il contesto socio-economico sta cambiando rapidamente. Mai come oggi è importante essere aggiornati sulle politiche che le **Nazioni Unite**, l'Europa e l'Italia stanno vagliando, comprendere l'impatto della pandemia sul mondo del lavoro, come il PNRR inciderà sulla transizione ecologica e digitale e sull'occupazione e come le organizzazioni possano affrontare questi cambiamenti per uno sviluppo sostenibile.

Future of Workers propone uscite periodiche per affrontare temi di rilievo a partire dalle ricerche, competenze ed esperienze della Fondazione Giacomo Brodolini.

Questo numero è dedicato a **Le sfide dell'intelligenza artificiale per una società più inclusiva.**



LA VOCE DI...

"Sento il dovere di dirvi che, in un caso come questo, il Ministro del Lavoro non pretende di collocarsi al di sopra delle parti, ma che sta con tutto il cuore da una sola parte: dalla vostra parte".

Giacomo Brodolini
al presidio dell'Apollon
in Via Veneto
(Roma, 31 dicembre 1968)



INDICE

L'ESPERTA DELL'ARGOMENTO.....	5
IN SINTESI.....	6
GLOSSARIO.....	7
TEMA.....	10
TREND.....	13
POLITICHE E PRATICHE IN ITALIA E IN EUROPA.....	26
QUADRO NORMATIVO.....	32
RISORSE PER APPROFONDIRE.....	37
ALCUNI NOSTRI PROGETTI.....	39

L'ESPERTA DELL'ARGOMENTO

Questo numero è a cura di

Barbara del Micheli

Coordinatrice dell'Unità Giustizia Sociale della Fondazione Giacomo Brodolini

Interessata alla realizzazione delle pari opportunità di genere sin dalla sua tesi di laurea, da oltre vent'anni lavora come ricercatrice, formatrice e project manager in iniziative che affrontano questo tema in vari ambiti: dalle politiche sociali, alle dinamiche organizzative e alla ricerca scientifica. Già freelance e imprenditrice, dal 2008 lavora alla Fondazione Giacomo Brodolini e dal 2019 coordina il gruppo che si occupa di Social Justice. Nel 2012 ha dato vita al primo master in Italia su Gender equality, diversity e inclusion, giunto nel 2024 alla sua sedicesima edizione. Dal 2010 fa parte della redazione di *ingenerare.it*, dove scrive prevalentemente di inclusione, nuove tecnologie e cambiamento organizzativo. Nel 2020 ha conseguito il dottorato in Lavoro, sviluppo e organizzazione presso l'Università degli studi di Modena e Reggio Emilia, con una tesi sulla relazione tra lo spazio e l'organizzazione alla luce dei cambiamenti introdotti dai processi di digitalizzazione. Dal 2012 lavora alla definizione di metodologie per la realizzazione di Gender Equality Plans (GEP) nelle organizzazioni che si occupano di ricerca scientifica. In qualità di consulente ha seguito la definizione dei GEP di Institute Pasteur, Fondazione regionale ricerca biomedica Regione Lombardia, Toscana Lifescience Foundation, Ospedale San Raffaele, Università Vita Salute. Recentemente si è occupata degli sviluppi dell'intelligenza artificiale e dell'impatto che possono avere in termini di rispetto dei diritti fondamentali e rischio di rafforzamento delle disuguaglianze.



IN SINTESI

TEMA

L'intelligenza artificiale



L'intelligenza artificiale (AI) è destinata ad influire sempre di più sulle nostre vite. Eppure le informazioni e i dati che ci restituisce sono meno neutri di quanto potremmo pensare e rischiano di esasperare le disuguaglianze esistenti nella società. Poiché le AI imparano processando i dati che esistono in rete, dati che inevitabilmente riflettono la visione e gli interessi di chi ha contribuito ad alimentare la rete stessa, l'intelligenza artificiale rischia di operare delle scelte condizionate e di amplificare le disuguaglianze esistenti. Le soluzioni non sono soltanto tecnologiche e richiedono tempi lunghi, nel frattempo è importante creare consapevolezza sul funzionamento dell'AI e lavorare ad un suo sviluppo ed utilizzo più trasparente, equo e regolamentato.

PRATICHE E POLITICHE

Strategie italiane e europee



La strategia europea in materia di intelligenza artificiale è contenuta nel *Libro Bianco sull'intelligenza artificiale*, che delinea i principi fondamentali per un quadro normativo per l'AI in Europa evidenziando come quest'ultimo si debba fondare sui valori fondamentali dell'UE, sul rispetto dei diritti umani e sulla tutela rispetto ai rischi di discriminazione che sono associati all'utilizzo di questa nuova tecnologia. A livello nazionale, il *Programma Strategico Intelligenza Artificiale 2022-2024* non prevede ancora un focus specifico sulla lotta alle discriminazioni e sulla promozione di un'AI più inclusiva, al netto di un'attenzione al superamento del divario di genere nei percorsi formativi previsti.

TREND

Alcuni numeri



IN EUROPA

- 78%** consapevolezza delle imprese rispetto all'importanza dell'AI
- 42%** delle imprese ha adottato almeno una tecnologia di AI, il 25% almeno due
- 18%** prevede di adottare l'intelligenza artificiale nei prossimi anni
- 40%** delle persone in Europa è consapevole dell'esistenza di rischi di discriminazione nell'utilizzo dell'AI

IN ITALIA

- 500** milioni di euro il valore del mercato dell'AI
- 32%** crescita del mercato nel 2022
- 61%** delle grandi imprese italiane ha avviato almeno un progetto di AI
- 37%** delle persone in Italia è consapevole dell'esistenza di rischi di discriminazione nell'utilizzo dell'AI

QUADRO NORMATIVO

Le norme



Il quadro normativo per la regolamentazione dell'AI è in continua evoluzione ed è fortemente condizionato dalla prospettiva europea. I suoi pilastri sono



il GDPR



la Dichiarazione europea sui diritti e i principi per il decennio digitale (2023/C 23/ 01)



la Proposta di regolamento comunitario sull'intelligenza artificiale approvata dal Parlamento europeo il 13 marzo 2024

GLOSSARIO

Di cosa parliamo quando parliamo di...

Algoritmo: l'algoritmo è un procedimento per risolvere un problema. Nel campo dell'informatica, si traduce in una sequenza di operazioni elementari, dette istruzioni, eseguibili da un computer. Gli algoritmi che muovono l'intelligenza artificiale sono composti da operazioni che per un computer sono elementari; la complessità di questi procedimenti deriva da come le operazioni sono combinate tra loro. La gestione degli algoritmi, che passa per l'assegnazione di compiti alle persone che lavorano, è tipicamente associata alla *gig economy*, un fenomeno ancora in gran parte non regolato che pone sfide crescenti a lavoratori e lavoratrici che operano attraverso le piattaforme.

ChatGPT: si tratta di un chatbot che sfrutta l'intelligenza artificiale generativa e l'apprendimento automatico. L'acronimo sta per *Generative Pre-trained Transformer* e indica uno strumento di elaborazione del linguaggio naturale che utilizza algoritmi avanzati di apprendimento automatico per generare ri-

sposte simili a quelle umane all'interno di un discorso. È in grado di generare testi credibili e scritti in linguaggio naturale sulla base di input puntuali, consentendo di scrivere mail, articoli, saggi su ogni argomento, canzoni, poesie, sceneggiature, ecc. È stato lanciato a fine novembre 2022 dalla società di ricerca statunitense OpenAI.

Bias: sono distorsioni dei fenomeni osservati, figlie di un pregiudizio, che incidono negativamente, pregiudicandone, appunto, la spiegazione corretta.

Computer vision: tecnologia emergente che consente ai computer di interpretare e analizzare i dati visivi che vengono raccolti dal mondo circostante. Implica lo sviluppo di algoritmi e tecnologie che consentono alle macchine di capire video e immagini, così come fanno gli esseri umani, utilizzando telecamere che permettono di ricostruire un modello di realtà in 3D partendo da immagini in 2D.



Dall-E 2: algoritmo neurale multimodale di intelligenza artificiale capace di generare immagini a partire da descrizioni testuali, sviluppato da OpenAI. Combina gli input provenienti dall'ambiente esterno per produrre informazioni precise e dettagliate all'esito di un dinamico processo di costante apprendimento, conforme alle modalità tecniche del *machine learning*.

Data labeling: processi di etichettatura dei dati grezzi (ad esempio immagini, file di testo, video, ecc.) finalizzati a creare *dataset di training* per gli algoritmi di apprendimento automatico che consentono di addestrare correttamente i modelli di *machine learning*.

Deep learning: tecnica dell'intelligenza artificiale in cui reti neurali artificiali multi-livello e algoritmi basati sulla struttura e il funzionamento del cervello umano apprendono da una grande quantità di dati per sviluppare *pattern decisionali*. Può essere considerato una branca del *machine learning*.

Diversity, equity e inclusion: diversità, equità e inclusione sono strutture organizzative che cercano di promuovere "il trattamento equo e la piena partecipazione di tutte le persone", in particolare dei gruppi "che storicamente sono stati sottorappresentati o soggetti a discriminazione" sulla base dell'identità o della disabilità.

Intelligenza artificiale: tecnologia che facilita e supporta l'applicazione di molte altre tecnologie, mostrando un comportamento intelligente che è funzionale ad analizzare un dato ambiente e a compiere azioni, con un certo grado di autonomia, per raggiungere obiettivi specifici. Per fare ciò, l'intelligenza artificiale utilizza strumenti di *machine learning* e *deep learning* per estrarre informazioni da un'enorme quantità di dati e generare nuovo valore basato su modelli costruiti attraverso quelle informazioni.

Intelligent data processing: tecnologia autonoma avanzata per l'estrazione e l'elaborazione di informazioni da diverse tipologie di dati, strutturati o meno. L'obiettivo principale di questa tecnologia è quello di eseguire un'estrazione veloce delle informazioni, senza però compromettere la precisione del processo.

Intelligent object: oggetti in grado di eseguire azioni e prendere decisioni senza richiedere l'intervento umano, interagendo con l'ambiente circostante mediante l'utilizzo di sensori (ad esempio termometri, videocamere, microfoni, rilevatori di inquinamento, ecc.) e attuatori (ad esempio apertura/chiusura porte/finestre, attivazione elettrodomestici e impianti, ecc.).

Internet of Things (IoT): termine che si utilizza per identificare un insieme di dispositivi fisici e di sensori che, applicati a macchinari o a strumenti produttivi, ricevono e scambiano dati attraverso una rete senza fili capace di connettersi a internet con differenti scopi, che includono il monitoraggio dei processi produttivi e delle *performance* delle persone.

AI del linguaggio: soluzioni di elaborazione del linguaggio basate su modelli linguistici che hanno portato le macchine ad acquisire la capacità di comprendere e generare linguaggio in base agli input e ai dati ricevuti, con finalità che possono variare dalla comprensione del contenuto, alla traduzione, fino alla produzione autonoma di testi, post, commenti, notizie, ecc.

Machine learning: applicazione dell'intelligenza artificiale che conferisce alle macchine la capacità di imparare automaticamente e di migliorare con l'esperienza senza essere esplicitamente programmate. Nel *machine learning* i programmi diventano sempre più intelligenti col tempo, mediante un processo di apprendimento che può essere supervisionato o non supervisionato dagli esseri umani.

Natural language processing:, con quest'espressione, tradotta in italiano come "elaborazione del linguaggio naturale", si intendono algoritmi di intelligenza artificiale in grado di analizzare, rappresentare e quindi comprendere il linguaggio naturale, cioè quello utilizzato dalle comunità umane per comunicare. Le finalità possono variare dalla comprensione del contenuto, alla traduzione, fino alla produzione di testo in modo autonomo a partire da dati o documenti forniti come input.

Reti neurali: prendono ispirazione direttamente dalla biologia umana. Esattamente come il nostro cervello è formato da neuroni connessi attraverso delle sinapsi, una rete neurale artificiale è composta da diversi neuroni artificiali collegati fra loro per lavorare insieme. Per poter risolvere un determinato problema, la rete neurale deve essere addestrata. In pratica, questo processo consiste nell'apprendere le connessioni fra i neuroni e i loro pesi, cioè la loro importanza, sulla base di un insieme di coppie input-output di addestramento. Al momento, anche le reti neurali artificiali più avanzate sono lontanissime dalle prestazioni di un cervello biologico.

Robotica: branca della cibernetica che studia la costruzione e le possibili applicazioni pratiche dei robot e, in particolare, lo sviluppo di sistemi di automazione industriali molto avan-

zati, dove i robot sostituiscono gli esseri umani in operazioni pericolose, faticose o ripetitive. La robotica è una delle branche dell'ingegneria in cui l'applicazione dell'intelligenza artificiale è più evidente.

Stereotipi: idee preconcepite e modelli di condotta socialmente costruiti che definiscono in modo generalizzato e semplificato i ruoli da ricoprire, le qualità, le attitudini possedute, gli ambiti di azione privilegiati, le manifestazioni della propria appartenenza di genere, ecc. Contribuiscono a giustificare e a mantenere relazioni di potere impari tra le categorie che ne sono oggetto, limitandone le possibilità di lavoro, carriera e affermazione sociale.

Web scraping: dall'inglese "to scrape" (grattare, raschiare), è una particolare tecnica di *crawling*. Un *crawler* è un software che ha lo scopo di raccogliere tutte le informazioni necessarie per indicizzare in modo automatico le pagine di un sito, analizzare i collegamenti ipertestuali, trovare associazioni fra termini di ricerca e classificarli. È largamente usato da tutti i motori di ricerca per offrire agli utenti risultati sempre aggiornati. Si tratta di un sistema in grado di estrapolare una grande varietà di informazioni: dati di contatto, indirizzi di posta elettronica, numeri di telefono, così come singoli termini di ricerca o url.



TEMA

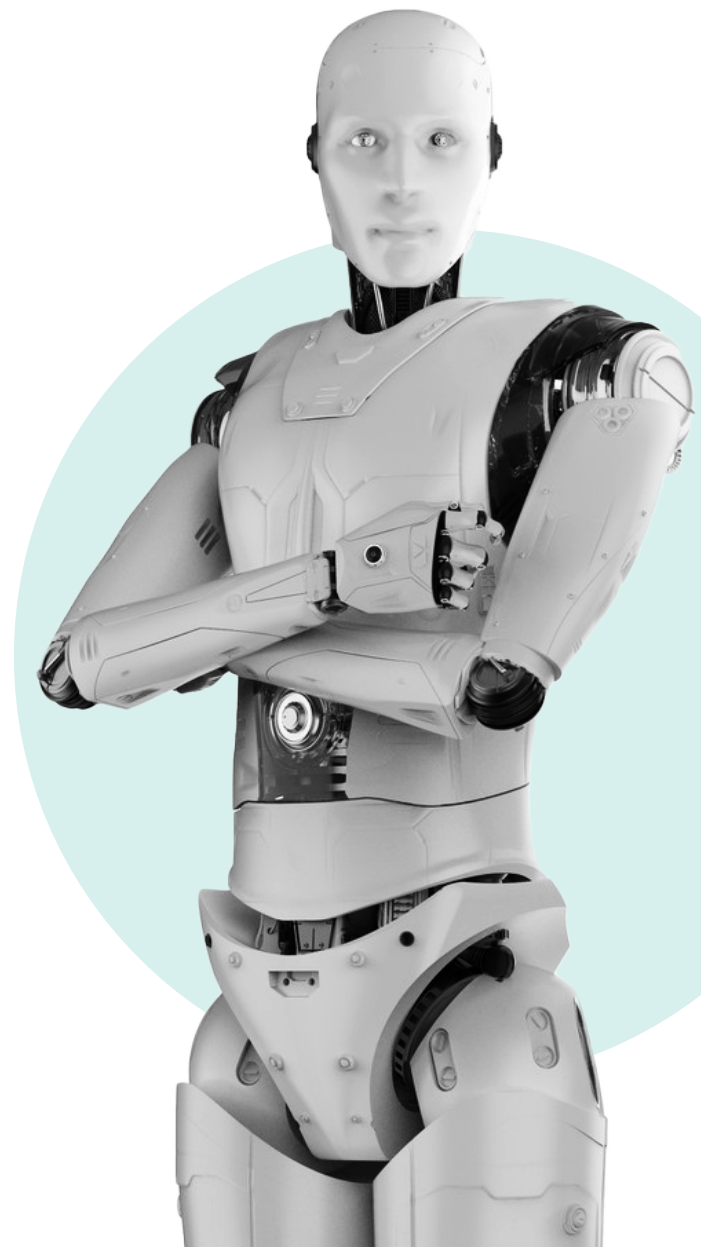
COSA INTENDIAMO PER INTELLIGENZA ARTIFICIALE

Sentiamo parlare sempre più spesso di *intelligenza artificiale* (in inglese *Artificial Intelligence*, abbreviato in AI), ma a cosa si fa riferimento esattamente? Secondo il gruppo di [High Level Experts sull'Intelligenza Artificiale dell'Unione europea](#), il termine si riferisce a "sistemi che mostrano un comportamento intelligente analizzando il loro ambiente e intraprendendo azioni - con un certo grado di autonomia - per raggiungere obiettivi specifici. I sistemi basati sull'intelligenza artificiale possono essere puramente basati su software, agendo nel mondo virtuale (ad esempio assistenti vocali, software di analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale); oppure, l'intelligenza artificiale può essere incorporata in dispositivi hardware (ad esempio robot avanzati, automobili autonome, droni o applicazioni Internet of Things). Si tratta, in altri termini, di simulare i processi dell'intelli-

genza umana attraverso la creazione e l'applicazione di algoritmi integrati in un ambiente di calcolo dinamico. Allo stato attuale di sviluppo, dietro ogni sistema di AI c'è il lavoro di una (o più) intelligenze umane, anche se questo lavoro non è necessariamente visibile.

COME FUNZIONA L'INTELLIGENZA ARTIFICIALE

Per far funzionare l'AI sono necessarie tre componenti chiave: sistemi di calcolo, dati e sistemi per la gestione dei dati e algoritmi avanzati, ovvero istruzioni passo dopo passo su come completare un compito specifico per trasformare i dati in informazioni utilizzabili. In questi ultimi anni stiamo assistendo a uno sviluppo esponenziale dell'intelligenza artificiale, perché si stanno verificando le condizioni ottimali: abbiamo a disposizione una grandissima e crescente quantità di dati e una maggiore potenza di elaborazione a costi relativamente più contenuti.



IN CHE MODO L'AI INFLUISCE SULLA NOSTRA VITA QUOTIDIANA

Sono sempre più numerosi i contesti della nostra vita quotidiana in cui troviamo applicazioni dell'intelligenza artificiale, e negli anni a venire sono destinati ad aumentare ulteriormente. Al momento le applicazioni sono in grado di risolvere problemi specifici in campi ben definiti, la cosiddetta "intelligenza artificiale debole", ma non siamo lontani dall'avvento della cosiddetta "intelligenza artificiale generale o forte", in cui i software saranno in grado di agire in contesti meno definiti con un approccio più globale.

PERCHÉ L'INTELLIGENZA ARTIFICIALE PUÒ ACCENTUARE LE DISCRIMINAZIONI E LE DISUGUAGLIANZE

Come abbiamo visto, il funzionamento dell'intelligenza artificiale si basa sulla disponibilità e sulla capacità di processare un numero sempre crescente di informazioni e dati; è sulla base dei dati disponibili che le macchine intelligenti costruiscono relazioni, formulano ipotesi e prendono decisioni. Tuttavia, le informazioni e i dati disponibili sono condizionate da una visione disuguale e stereotipata della realtà: mentre alcuni gruppi di persone sono

sovrarappresentati, altri sono del tutto assenti, e ad alcune tipologie di persone vengono attribuite determinate caratteristiche – positive o negative – più spesso che ad altre. E, poiché la macchina processa le informazioni a cui ha accesso, tende a riprodurre ed esasperare le disuguaglianze esistenti.

I FATTORI DI RISCHIO

In questo momento di sviluppo esponenziale dell'intelligenza artificiale, si pone con forza il tema della tutela degli individui, sia rispetto a sistemi di controllo potenzialmente molto invasivi, sia rispetto a uno sviluppo altrettanto esponenziale delle disuguaglianze.

Se gli algoritmi governeranno ambiti sempre più ampi della nostra vita, compreso l'accesso ai servizi pubblici essenziali, all'istruzione e al mondo del lavoro, è molto importante che sia gli algoritmi che i dati che questi ultimi processano siano condizionati il meno possibile da stereotipi e pregiudizi (*bias*), e che tendano a limitare invece che a esasperare esclusioni e disuguaglianze.

Esempi di potenziali problemi in relazione ai diritti fondamentali legati all'uso delle tecnologie basate sull'intelligenza artificiale emergono ormai sempre più spesso.

Ad esempio, alcuni [studi](#) hanno evidenziato come un algoritmo utilizzato per reclutare risorse umane preferisca generalmente gli uomini rispetto alle donne; un chatbot online è diventato "razzista" nel giro di un paio d'ore, imparando dagli umani con cui interagisce; le traduzioni automatiche hanno mostrato pregiudizi di genere, con la tendenza a utilizzare preferibilmente il maschile nelle lingue di arrivo (il cosiddetto "maschile sovraesteso"); i [sistemi di riconoscimento facciale](#) rilevano con esattezza il genere per gli uomini bianchi, ma non per le donne nere – perché queste ultime sono meno presenti nelle banche dati da cui i sistemi apprendono, e quindi la possibilità di esercitarsi alla lettura dei dettagli dei visi è minore. Emerge poi il tema della catalogazione e dell'indicizzazione delle informazioni e della conoscenza, un problema non nuovo ma che l'alta mole di dati disponibili tende a esasperare.



IL NOSTRO GRADO DI CONSAPEVOLEZZA COME CITTADINE, CITTADINI E UTENTI

La nostra consapevolezza nei riguardi dell'intelligenza artificiale è in generale molto bassa, con atteggiamenti che vanno dallo scetticismo (e la paura) per gli effetti invasivi che un suo sviluppo potrà avere sulle nostre vite, fino a un'ottimistica presunzione di obiettività legata a una fiducia totale nella tecnologia.

Pensiamo ad esempio a cosa accade quando, per acquisire informazioni su qualcosa che non conosciamo, facciamo una ricerca su un motore di ricerca online: la maggior parte di noi si ferma ai risultati contenuti nella prima pagina, basando la costruzione di un'opinione sull'argomento sulle informazioni che ci vengono presentate. Eppure, spesso questi risultati escludono alcuni gruppi di persone o le rappresentano in modo costantemente negativo.

Nonostante alcuni casi clamorosi abbiano costretto alcuni motori di ricerca ad ammettere che la propria attività produce risultati condizionati dagli stereotipi e non sempre del tutto attendibili (come la [gaffe razzista](#) del 2015, in cui due persone nere sono state etichettate dal software Google Foto come gorilla), la reazione delle multinazionali della conoscenza è stata limitarsi a scusarsi e dichiarare di non poter intervenire per ragioni legate alla tecnologia e alla necessità di difendere il diritto alla libertà di espressione.

La disuguaglianza nella disponibilità di informazione pone anche il problema, che esiste anche per altre applicazioni che utilizziamo sul web, dell'utilizzo solo apparentemente gratuito per l'utente finale. Quando ci sembra di usufruire gratuitamente di un servizio, in realtà paghiamo con una contropartita di informazioni – i nostri gusti, le nostre ricerche, le nostre curiosità inconfessabili – talmente ricca da consentire un livello di profilazione e conoscenza che spesso supera l'intimità e la condivisione che abbiamo con le persone che vivono con noi.

MISURE E OPPORTUNITÀ PER TRASFORMARE I RISCHI IN FATTORI DI INCLUSIONE

Quello di stabilire dei limiti e governare l'intelligenza artificiale è uno dei temi più caldi del dibattito attuale. Da un lato, ci sono i fautori dell'assenza di regolamentazione a favore di una crescita incondizionata delle potenzialità di strumenti che potrebbero migliorare, da molti punti di vista, la vita quotidiana di tutti e tutte noi. Dall'altro, c'è chi si pone con forza il problema di garantire un accesso ampio a queste potenzialità – alcune delle quali sono già esistenti ma accessibili soltanto alla parte più ricca della popolazione mondiale – nel rispetto dei diritti fondamentali e senza dare adito a discriminazioni.

Le misure proposte vanno dall'autoregolamentazione delle multinazionali della conoscenza alla definizione di un sistema sovranazionale di regole stringenti che tutelino le singole persone, alla richiesta di trasparenza sui software utilizzati e sulle loro finalità e alla necessità di far crescere il livello di consapevolezza di cittadini e cittadine per un uso consapevole degli strumenti di AI, fino al maggiore coinvolgimento dei gruppi sotto-rappresentati nella programmazione dei software che regolano l'intelligenza artificiale. Superare gli stereotipi nei *big data* è una sfida di proporzioni enormi, ma che potrebbe contribuire a ridurre le disuguaglianze anche nel mondo reale.

Allo stato attuale dello sviluppo tecnologico, con alcune parti del processo che non sono completamente trasparenti e un'opacità diffusa rispetto alle caratteristiche dei *dataset* su cui l'AI impara e funziona, diventa importante che le persone siano sempre più consapevoli del fatto che questo condizionamento esiste e che può essere tecnicamente difficile da eliminare; che l'intelligenza artificiale è una tecnologia, e che dobbiamo imparare a utilizzarla in quanto tale.

TREND

DA DOVE NASCE E COME FUNZIONA L'INTELLIGENZA ARTIFICIALE

L'espressione "intelligenza artificiale" non è nuova nasce nel 1956, in occasione della conferenza [Dartmouth Summer Research Project on Artificial Intelligence](#) a cui partecipò l'informatico statunitense John Mac Carthy, che è considerato l'inventore moderno di quest'espressione. Nell'ambito della conferenza, si iniziò a parlare di [machine learning](#), [deep learning](#), analisi predittiva e scienza dei dati, che sono gli elementi su cui si basa lo sviluppo dell'intelligenza artificiale.

Eppure, la storia dell'intelligenza artificiale comincia ancora prima: ai suoi albori troviamo [Ada Lovelace](#)¹, matematica londinese e prima programmatrice della storia, a cui va il merito di aver sviluppato il primo software, nel 1843, e di aver capito che le macchine computazionali

avrebbero potuto trascendere il campo dei soli numeri elaborando qualunque tipo di informazione, dalle parole, alle immagini, alla musica. Fu Ada Lovelace, infatti, a descrivere il primo algoritmo come una serie finita di istruzioni per risolvere un problema, che doveva permettere alla macchina analitica ideata da Babbage di calcolare un elemento della serie dei numeri di Bernoulli senza dover calcolare i suoi precedenti. Nonostante la precarietà dei mezzi tecnologici disponibili all'epoca, fu Lovelace a gettare le basi della moderna informatica e a iniziare a porsi domande filosofiche ed etiche intorno alle macchine e alle loro potenzialità, inclusa la questione ancora oggi cruciale se le macchine possano raggiungere, e magari superare, il livello di elaborazione del pensiero umano.



¹ L'[istituto](#) che porta il suo nome svolge ancora oggi un importantissimo lavoro di ricerca su questi temi.



Se la risposta di Lovelace, formulata sulla base delle informazioni a lei disponibili, è che le intelligenze artificiali non avrebbero mai potuto sorprendere i propri programmatori, sarà Alan Turing a mettere a punto, nel 1950, il famoso test per determinare l'intelligenza di una macchina, a partire dalle intuizioni di Lovelace. Secondo Turing, se in un dialogo una persona non è in grado di determinare con successo se il suo interlocutore sia un altro essere umano o un computer, si è al cospetto di una macchina dal comportamento intelligente. Chi tra noi che ha avuto almeno una conversazione con un chatbot sa come oggi la maggior parte sia perfettamente in grado di superare il test di Turing, mentre non tutti superano [i nuovi test sviluppati per misurare l'AI](#).

A oggi, il funzionamento di un sistema di intelligenza artificiale viene definito grazie a [quattro differenti livelli funzionali](#), in grado di esaurire le operazioni che è chiamato a effettuare:

- **comprensione:** la capacità di apprendere e simulare la correlazione tra i dati e gli eventi, attraverso cui l'AI può, ad esempio, riconoscere testi, immagini, video, audio e voce per elaborare informazioni specifiche sulla base di una specifica richiesta.
- **Ragionamento:** la capacità dei sistemi AI di collegare in maniera logica e in totale autonomia i dati raccolti.
- **Apprendimento:** la capacità di sistemi di AI di analizzare l'input dei dati per restituire un output corretto. È il caso dei sistemi di

apprendimento automatico ([machine learning](#)) che utilizzano tecniche specifiche per apprendere da un contesto informativo, per svolgere determinate funzioni.

- **Interazione:** si tratta di sistemi di *human machine interaction* (HMI) dove l'AI esercita un ruolo fondamentale nella sua interazione con l'essere umano. Uno degli esempi più ricorrenti di questo livello funzionale è dato dal [natural language processing](#) (NLP), l'insieme di tecnologie che consente di creare una relazione verbale tra l'uomo e la macchina, sfruttando il linguaggio naturale, come avviene nel caso dei chatbot più evoluti.

LA DIFFUSIONE DELL'INTELLIGENZA ARTIFICIALE IN EUROPA OGGI

Secondo la [European Enterprise Survey](#), nel 2020 la consapevolezza delle imprese sull'AI era elevata in tutta l'Unione europea (78%). Quattro imprese su dieci (42%) avevano adottato almeno una tecnologia AI, il 25% ne aveva adottate almeno due, mentre il 18% prevedeva di adottare l'intelligenza artificiale nei successivi due anni; il 40% non aveva adottato l'intelligenza artificiale né aveva intenzione di farlo. L'adozione a livello di ciascuna tecnologia è ancora relativamente basso: dal 3% per l'analisi del sentiment al 13% per il rilevamento di anomalie e ottimizzazione nei processi e nelle attrezzature. La strategia di approvvigionamento più comune è esterna: il 59% delle imprese dell'UE utilizza software o sistemi di AI pronti all'uso. Per le imprese, i principali ostacoli interni all'adozione dell'AI sono le difficoltà nell'assumere nuovo personale con le giuste competenze (57%), il costo di adozione (52%) e per l'adattamento dei processi operativi (49%). Ridurre l'incertezza può essere vantaggioso, poiché le imprese si assumono la responsabilità per potenziali danni (33%), standardizzazione dei dati (33%) e gli ostacoli normativi (29%) rappresentano le principali sfide esterne all'adozione dell'AI.

L'AI IN ITALIA ?

Secondo l'Osservatorio Digital Innovation Intelligenza Artificiale del Politecnico di Milano, in Italia il mercato dell'AI nel 2022 ha raggiunto 500 milioni di euro, con una crescita di ben il 32% in un solo anno, di cui il 73% commissionato da imprese italiane (365 milioni di euro) e il 27% rappresentato da export di progetti (135 milioni di euro).

Il 2022 è stato un anno da record, caratterizzato dai continui progressi nelle capacità delle macchine e dagli exploit di [Dall-E2](#) e [ChatGPT](#), che in poche settimane hanno coinvolto decine di milioni di utenti e mostrato al grande pubblico le potenzialità di questa tecnologia.

Sempre secondo l'osservatorio, oggi il 61% delle grandi imprese italiane ha già avviato almeno un progetto di AI, con un aumento di 10 punti percentuali rispetto a cinque anni fa. E tra queste, il 42% ne ha più di uno operativo. Tra le piccole e medie imprese, invece, il 15% ha almeno un progetto di AI avviato (nel 2021 era il 6%), quasi sempre uno solo, ma una su tre ha in programma di avviarne di nuovi nei prossimi due anni.



I PRINCIPALI AMBITI DI APPLICAZIONE

Secondo l'Osservatorio digitale del Politecnico di Milano, le soluzioni di intelligenza artificiale attualmente esistenti possono essere ricondotte alle tipologie presentate nella tabella che segue.



Intelligent data processing: si tratta di una tecnologia per estrapolare informazioni tramite algoritmi in grado di analizzare dati specifici e compiere azioni in conseguenza. Viene ad esempio utilizzata per l'analisi predittiva (analisi dei dati per fornire previsioni sugli andamenti futuri) e il rilevamento di frodi (identificazione di elementi non conformi rispetto a un modello previsto).



Virtual assistant/Chatbot: i chatbot utilizzano tecnologie di linguaggio naturale per eseguire azioni ed erogare servizi per gli utenti, sulla base di un'interazione vocale o testuale. Le tecnologie basate sul *natural language processing* consentono alle applicazioni chatbot più evolute di comprendere anche il tono della conversazione, oltre a memorizzare le informazioni raccolte, implementando ad esempio i dataset dei sistemi di *customer relationship management* (CRM) connessi.



Recommendation: sulla base di informazioni sul comportamento degli utenti, acquisite in forma di diretta o indiretta, l'AI è in grado di integrare i sistemi di vendita con applicazioni che danno suggerimento al consumatore per guidarlo verso la finalizzazione di un acquisto. Consentono pertanto di aumentare le conversioni e ottimizzare l'efficienza del cosiddetto *customer journey*, influenzando il processo decisionale del cliente



Image processing: le tecnologie di visione computerizzata basate sul riconoscimento delle immagini consentono di riconoscere in maniera automatica oggetti, persone e animali, per supportare moltissime applicazioni che vanno dalla videosorveglianza al rilevamento dei pattern anomali nel controllo qualità industriale.



Autonomous vehicle: le AI consentono di supportare i sistemi di guida autonoma dei veicoli. La guida autonoma è diventata molto popolare in ambito *automotive*, ma riguarda anche altri mezzi di trasporto autoguidati: supporta infatti la navigazione marittima, fluviale e aerea.



Intelligent object: o *smart object*. Gli oggetti intelligenti classificano i dispositivi in grado di effettuare operazioni senza l'input diretto dell'utente umano. I sistemi AI integrati, generalmente associati a sensori intelligenti, consentono agli oggetti intelligenti di prendere decisioni sulla base delle condizioni dell'ambiente circostante, attraverso reazioni a un determinato evento.



Language processing: comprende tutte le applicazioni basate sulla comprensione del testo e della comunicazione verbale, attraverso le tecnologie di elaborazione del linguaggio naturale (NLP).



Autonomous robot: grazie alle istruzioni di sistemi basati su tecnologie AI, i robot possono muoversi e agire senza che vi sia un utente a impartirgli un comando diretto, potendo contare sulla capacità di riconoscere e interagire con l'ambiente circostante. La robotica autonoma è sempre più diffusa sia nei contesti di manifattura industriale che nella logistica e nelle applicazioni civili/domestiche.

Nella pratica, questo implica una varietà sempre crescente di applicazioni che hanno un impatto sulle modalità di lavoro nelle imprese, sulle relazioni con la pubblica amministrazione, sulla nostra vita quotidiana come cittadine e cittadini, come illustrato dalla tabella che segue.

AMBITO DI APPLICAZIONE	TIPO DI APPLICAZIONE
Agricoltura	→ Aumento dell'efficienza dei sistemi di produzione mediante la manutenzione predittiva, contribuendo alla mitigazione e all'adattamento ai cambiamenti climatici
Grande distribuzione	→ Analisi predittiva dei consumi e gestione di scorte e distribuzione
Sanità	→ Supporto alla diagnostica, alla cura, alla prevenzione e alla ricerca medica
Industria manifatturiera 4.0	→ Sensori sugli impianti e possibilità di prevenire possibili guasti
Selezione del personale	→ Scrematura dei cv sulla base delle indicazioni del cliente
E-commerce	→ Possibilità di creare esperienze di acquisto su misura per l'utente
Marketing	→ <i>Sentiment analysis</i> delle tracce comportamentali sul web e definizione di campagne che intercettino desideri della clientela
Assistenza clienti e customer care	→ Gestione FAQ con chatbot che interagiscono parlando con gli utenti
Giornalismo e servizi editoriali	→ Produzione testi originali, sistemi di traduzione e editing
Automotive	→ Sviluppo veicoli a guida autonoma
Vigilanza antifrode	→ Rilevamento comportamenti anomali e attivazione procedure di blocco preventivo o avviso degli utenti
Sistemi GPS	→ Analisi predittiva dei dati e gestione del traffico
Domotica	→ Software unico per la gestione di elettromeccanici, impianti e sicurezza
Sistemi di controllo e sorveglianza	→ Riconoscimento biometrico e potenziamento dei sistemi di sicurezza
Entertainment e videogame	→ Esperienze culturali e di gioco sempre più realistiche, coinvolgenti e su misura per l'utente

Tra quelle presentate, anche le applicazioni più avanzate si basano al momento sulla cosiddetta **AI debole**, ovvero su sistemi progettati per risolvere problemi specifici, sebbene di complessità variabile. Nei fatti, queste applicazioni simulano alcune capacità umane di *problem solving*, senza pretendere di comprendere e replicare il funzionamento di tutto ciò che il cervello umano è capace di fare.

Nei vari ambiti di applicazione, l'obiettivo della macchina è risolvere il problema pratico nel migliore dei modi possibili. In quest'ottica, le applicazioni basate sull'intelligenza artificiale debole offrono soluzioni ottimali per eventi ripetitivi o percorsi consolidati, ad esempio le

domande frequenti a cui rispondono i chatbot dei servizi di supporto alla clientela, e funzionano bene in collaborazione con gli esseri umani perché suggeriscono quali decisioni prendere offrendo un maggior numero di informazioni a sostegno delle possibili scelte.

Siamo ancora lontani, invece, dalle applicazioni della cosiddetta **intelligenza artificiale forte o generale** (in inglese *Artificial General Intelligence*, AGI) ovvero da sistemi capaci di comportarsi in maniera del tutto autonoma a prescindere dal contesto e dal compito che viene loro assegnato. Questo approccio è radicalmente differente e supera anche la concezione di macchine intelligenti elaborata da Turing.

Si tratta di sviluppare una sorta di coscienza autonoma – generale, appunto – svincolata da contesti specifici e quindi capace di agire con successo in qualsiasi situazione, pensando in modo razionale e capace di apprendere dal contesto in cui si muove.

Anche se l'obiettivo è lungi dall'essere raggiunto, alcune applicazioni sviluppate in tempi più recenti offrono una visione chiara delle potenzialità. Si pensi a [GPT-4](#), [Generative Pre-trained Transformer 3](#) elaborato dal Laboratorio OpenAI di Microsoft, capace di elaborare testi e immagini in maniera formalmente indistinguibile rispetto a quanto farebbe un essere umano.



OPPORTUNITÀ E RISCHI DI DISCRIMINAZIONE

Nel volume di racconti pubblicato nel 2021 con il titolo evocativo *Intelligenza Artificiale 2041*, Kai-Fu Lee, ex presidente di Google China, e Chen Qiufan, astro nascente della fantascienza cinese, propongono dieci storie ambientate in un imminente futuro in cui declinano in applicazioni concrete le potenzialità non ancora espresse dell'intelligenza artificiale.

Il pregio principale dei racconti è di tener conto dell'impatto delle soluzioni tecnologiche sui vari scenari umani in cui si applicano e di spiegare a chi legge quale sia la tecnologia che rende possibile l'applicazione. Ciascun racconto è ambientato in un continente diverso e le condizioni sociali, economiche e culturali del contesto vengono incluse nella lettura delle potenzialità della tecnologia. Il tema principale non è l'AI, ma **l'umanità nelle sue differenze e nella sua interazione con le macchine**. I racconti ce ne mostrano le potenzialità ma anche i rischi, e il tema della diversità e dell'inclusione è declinato con cura in ogni narrazione.

Nel primo racconto, ad esempio, siamo in una Mumbai del 2041 dove è possibile sottoscrivere un programma assicurativo basato sul *deep learning*. Si tratta di un programma dinamico che si relaziona con la clientela tramite una serie di applicazioni che hanno lo scopo di migliorarne la vita. In altri termini, suggerisce comportamenti salutari e da evitare, sulla base

dello storico dei dati disponibili. Tra i comportamenti da evitare, per una ragazza adolescente della classe media cittadina, il sistema include la frequentazione di persone e luoghi della città collegati alle caste delle persone "intoccabili": anche se il sistema delle caste non esiste legalmente in India, l'AI "conosce la storia dei suoi utenti" e sa che queste persone sono considerate ai margini. Quando Nayana, la protagonista, si innamora di Sahej, la cui famiglia ha origini "intoccabili", lo smart watch con tracciamento GPS a cui la sua applicazione assicurativa è collegata aumenta il valore del premio assicurativo ogni volta che lei gli si avvicina. L'ottimizzazione della funzione obiettivo della app – ovvero rendere il premio assicurativo il più basso possibile – entra in conflitto con i sentimenti della ragazza.

La narrazione mette in evidenza alcuni temi chiave che rappresentano al tempo stesso un rischio e un'opportunità.

In primo luogo, troviamo il tema della **gestione della mole di dati** che ciascuno di noi condivide con le applicazioni che utilizza, che se, da un lato, aiuta senza dubbio ad avere accesso a servizi ed esperienze costruite su misura per noi, dall'altro è lecito chiedersi se ci siano informazioni che hanno diritto a essere tutelate, non condivise, mantenute private. In secondo luogo, c'è la necessità ma anche la difficoltà di definire **funzioni-obiettivo complesse**, che tengano conto delle differenze tra gli esseri

umani e della possibilità di includere nell'algoritmo il tema del benessere e della felicità. Se vogliamo che l'AI ne tenga conto, servono modelli per definire la felicità in termini matematici che considerino l'importanza dell'indeterminatezza e della sorpresa nella definizione della felicità individuale. Infine, c'è il **tema delle disuguaglianze, dell'iniquità e dei pregiudizi**. Da una parte, basando le sue decisioni soltanto sui dati e sull'ottimizzazione dei risultati, l'AI dovrebbe produrre decisioni più eque di quelle delle persone, che più facilmente possono essere influenzate da pregiudizi di vario tipo; dall'altro, l'AI impara da, e si basa su dati che potrebbero essere insufficienti a rappresentare in modo adeguato le differenze, in particolare demografiche etniche e di genere.

Quest'ultimo aspetto non riguarda soltanto lo sviluppo futuro dell'AI ma anche il presente: diverse aziende, che hanno iniziato ad utilizzare l'AI per le selezioni del personale, ad esempio [Amazon](#), si sono accorte che gli algoritmi penalizzano le donne e i gruppi di minoranza perché troppo poco presenti nei dati utilizzati per addestrarli – la macchina non li aveva visti e non li riconosceva.

La necessità di riflettere su come costruire un'intelligenza artificiale più inclusiva è estremamente rilevante anche alla luce delle questioni aperte sul suo sviluppo, riassunte nella tabella che segue:

QUESTIONE APERTA	DECLINAZIONE
Potere degli organismi statali e democrazia	L'intelligenza artificiale aumenta in modo esponenziale le possibilità di controllo da parte degli apparati statali. Questo elemento può contribuire a rendere più efficiente il funzionamento dei meccanismi e dei servizi, ma cosa può accadere nei regimi non democratici? Quali sono i limiti per le libertà individuali?
Rispetto dei diritti fondamentali e AI	L'analisi dell'impatto delle innovazioni legate all'AI si concentra prevalentemente sulla componente tecnologica, spesso senza considerare gli effetti sui diritti umani. Eppure, molte applicazioni dell'AI hanno un potenziale impatto su diversi diritti fondamentali, al di là del diritto alla privacy. Si pensi al diritto di non discriminazione, alla dignità, alla sicurezza e all'assistenza sociale, al diritto alla buona amministrazione e alla protezione dei consumatori.
Proprietà intellettuale e corretto utilizzo	Di chi sono i dati condivisi? Che diritti mantengono sui propri dati le singole persone? E a chi appartengono i contenuti condivisi sul web? È lecito che siano utilizzati dalle macchine per apprendere e poi produrre contenuti che si basano su di essi? Come si può regolamentare la proprietà dei dati e dei contenuti?
Monopolio di pochi big player	Attualmente lo sviluppo dell'AI è nelle mani di pochi player che, grazie all'ampia disponibilità di risorse, orientano gli sviluppi della ricerca in accordo con i propri interessi. Alcune applicazioni hanno però un impatto potenziale che solleva quesiti morali ed etici che non possono essere lasciati nelle mani di un ristretto gruppo di persone, per lo più appartenenti alla stessa estrazione sociale (uomini bianchi occidentali).
Accessibilità dei modelli di funzionamento	La concentrazione degli sviluppi dell'AI pone anche un problema di libera concorrenza e accesso al mercato. Le opportunità di utilizzo dell'AI saranno solo per le grandi imprese? Le PMI avranno la possibilità di internalizzare le competenze per gestire la tecnologia?
Conservazione dei dati	Dove finiscono i dati che condividiamo? Chi regola il loro utilizzo? Per quanto tempo rimangono negli archivi? Esiste un diritto all'oblio e come deve essere disciplinato? (si pensi alla recente normativa sull'oblio oncologico)
Trasparenza dei modelli	I modelli utilizzati dalle aziende che usano l'AI dovrebbero essere trasparenti, l'utente dovrebbe sapere quali modelli vengono utilizzati e per quali scopi (e decidere di conseguenza se condividere i propri dati)
Mancanza di accuratezza	L'AI fornisce risposte credibili ma spesso non accurate o false
Assenza di empatia dell'AI	L'AI non ha sentimenti come compassione ed empatia, non può fare in modo che le persone si sentano comprese, né mediare, laddove necessario, la rigidità dell'algoritmo. Quali conseguenze provoca l'utilizzo estensivo di chatbot sulle relazioni, con una popolazione mondiale sempre più anziana (che potrebbe avere bisogno di linguaggi diversi, spiegazioni aggiuntive, tempi differenti per interagire)?
Mancanza di creatività dell'AI	L'AI non può pensare strategicamente né applicare "il buon senso". Non riesce a cogliere la complessità di alcune situazioni che implicano sentimenti (cosa accade davanti a un imprevisto? Come viene gestita la necessità di compiere una scelta inusuale e non razionale?)
Eliminazione di pregiudizi e bias	L'AI attualmente lavora in un regime di polarizzazione della conoscenza, per cui si rende necessario raccogliere dati e creare conoscenza differenziale, inserire punti di vista diversi nella programmazione e dare maggiore variabilità ai dati. È importante valutare la qualità dei dati su cui l'AI apprende, soprattutto se il rating su questi ultimi è sempre più utilizzato per l'accesso a servizi e sistemi sociali

Facciamo un esempio per chiarire l'impatto potenziale delle discriminazioni a partire dall'ultimo dei punti della tabella.

Uno [studio](#) ha dimostrato che un algoritmo utilizzato negli Stati Uniti sottovaluta sistematicamente i bisogni sanitari dei e delle pazienti di pelle nera. Come *proxy* dei bisogni sanitari, l'algoritmo utilizza i costi sanitari sostenuti nelle annualità precedenti, ma quelli sostenuti dalle persone nere sono storicamente inferiori. Non perché fossero più sane, ma perché il razzismo sistemico ha impedito loro l'accesso alle cure. Questo esempio illustra come l'apprendimento automatico e l'intelligenza artificiale possono replicare e amplificare le disuguaglianze. La maggior parte degli algoritmi sfrutta le correlazioni grezze nei dati; eppure, queste correlazioni sono spesso sottoprodotti di relazioni sociali più salienti (nell'esempio dell'assistenza sanitaria, il non accesso alle cure è, per definizione, a costo zero in termini di acquisto servizi) o di eventi casuali che non si replicano.

L'intelligenza artificiale *dice quello che sa* (rielabora i dati di cui dispone) ma *non sa quello che dice* (non è in grado di valutare la veridicità o verosimiglianza di quanto sostenuto sulla base di parametri che non siano già nell'equazione).

Per identificare e mitigare le relazioni discriminatorie nei dati, abbiamo bisogno di modelli che catturino o tengano conto dei percorsi causali che a queste danno origine.

Dal momento che, tuttavia, questo processo è complesso e richiede tempi lunghi, è necessario che cittadini e cittadine siano sempre più consapevoli delle potenzialità e dei limiti della tecnologia per imparare a interagirvi in modo funzionale.

I primi tentativi di correzione dei *bias* degli algoritmi, infatti, producono risultati non ancora soddisfacenti: sembra che l'elemento più difficile da inserire nell'equazione sia la valutazione del contesto, la capacità di "situare" la risposta nello spazio e nel tempo. Si veda quello che sta succedendo con il primo rilascio dei [software di Google](#) per l'elaborazione di immagini. Un'operazione che la mente umana riesce a fare agilmente, tenendo contemporaneamente in considerazione una molteplicità di elementi per l'elaborazione.



FOCUS SU AI E DISCRIMINAZIONI DI GENERE

I dati disponibili sulle discriminazioni di genere in ambito AI consentono una lettura quasi esclusivamente binaria del fenomeno: negli studi disponibili il termine “genere” viene letto come distinzione tra maschile e femminile (senza considerare altri generi) e i dati disponibili riguardano quasi esclusivamente le donne. Il [rapporto](#) del 2022 dello European Institute for Gender Equality (EIGE) mette in evidenza alcuni elementi che caratterizzano la discriminazione delle donne negli ambiti legati all’AI.

- Le donne ricoprono in media meno del 20% dei posti di lavoro del settore in Europa.
 - La loro permanenza in un settore così fortemente maschile è a rischio: le donne tendono ad abbandonare più degli uomini – nonostante si tratti di posti di lavoro ben remunerati.
 - L’assenza di donne tra chi progetta e disegna l’AI si riflette nel permanere di stereotipi e *bias* nella progettazione.
 - Le donne sono oggetto di nuove forme di discriminazione e violenza (ad esempio i video porno falsi creati dall’AI utilizzando quasi esclusivamente immagini di persone a loro insaputa).
- I pregiudizi di genere nell’intelligenza artificiale si verificano in varie fasi:
 - » nel processo di sviluppo dell’algoritmo;
 - » nel training dei dataset;
 - » nel processo decisionale generato dall’intelligenza artificiale.
 - La tendenza a femminilizzare gli strumenti di intelligenza artificiale imita e rafforza le gerarchie strutturali e gli stereotipi nella società, che si basa su ruoli di genere pre-assegnati.
 - Agli assistenti virtuali domestici come Alexa di Amazon, Cortana di Microsoft e Siri di Apple sono state assegnate voci femminili predefinite (Apple e Google da allora hanno offerto alternative mirate alla “diversificazione” o alla “neutralità”). Tuttavia, come risulta evidente dal caso di Watson dell’IBM, che utilizzava una voce maschile mentre lavorava con i medici sulla cura del cancro, le voci maschili sono state preferite per compiti che implicavano insegnamento e istruzione, poiché erano percepite come “autoritarie e assertive”.
- Anche le forme addomesticate e femminilizzate di AI svolgono sempre più il “lavoro affettivo” che convenzionalmente ci si aspetta dalle donne. Inoltre, l’umanizzazione degli assistenti virtuali e dei robot può anche consentire la disumanizzazione e l’oggettificazione delle donne. Ad esempio, il robot umanoide Sophia è stato realizzato per sembrare “eccezionalmente attraente”, evocando un senso di “meccanico-erotismo”. Allo stesso modo, ai robot umanoidi vengono spesso attribuite “caratteristiche asiatiche o caucasiche” e sono “ipersessualizzati”. Di conseguenza, le varie forme di violenza di genere e di molestie che le donne e le ragazze affrontano sia nella sfera privata che in quella pubblica si riflettono anche nel modo in cui vengono trattate le forme femminilizzate di AI.

La sfida che si pone, in termini di costruzione di una società più giusta e meno discriminante, è come rendere l’AI un alleato nella lotta alle disuguaglianze, interrogandosi su come sia possibile incorporare sempre più equità e inclusione negli algoritmi.

IL GRADO DI CONSAPEVOLEZZA DI CITTADINI E CITTADINE EUROPEE

Il grafico che segue è tratto dalla [survey Eurobarometer](#), realizzata periodicamente per monitorare le opinioni di cittadine e cittadini europei su diverse questioni di carattere economico, politico e sociale.

Al quesito specifico sul tema della consapevolezza in merito ai potenziali effetti discriminatori dell'intelligenza artificiale, in media solo il 40% dei e delle rispondenti europee dichiara di essere consapevole dell'esistenza di un rischio discriminazione quando utilizza l'AI.

Se la consapevolezza varia di molto nei diversi paesi, l'Italia si allinea alla media europea, con un tasso di consapevolezza pari al 37%.



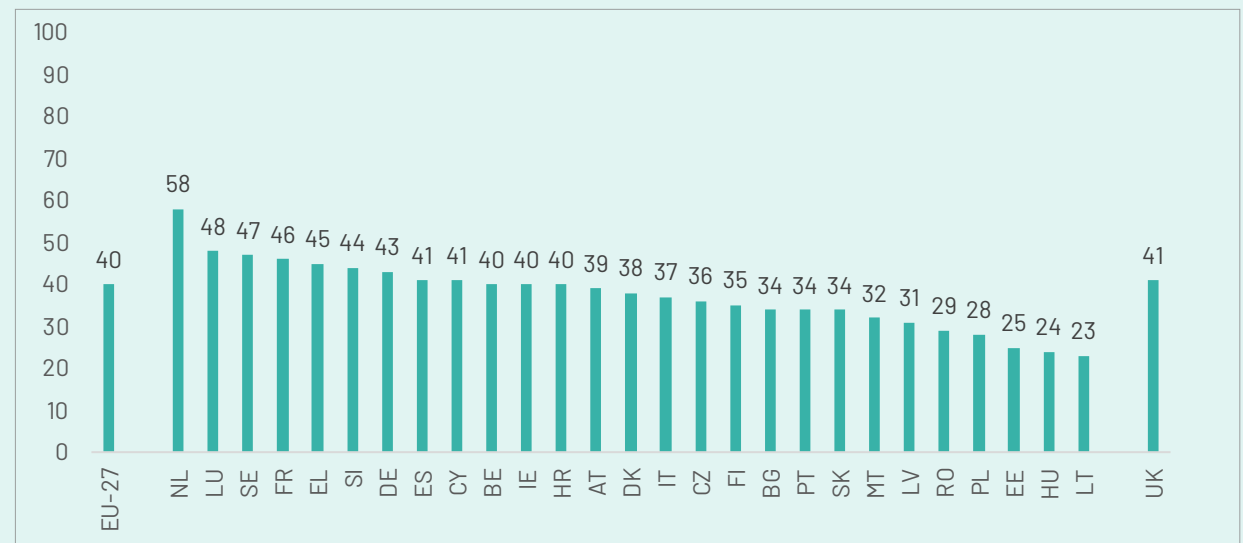
CONSAPEVOLEZZA DA PARTE DELLA POPOLAZIONE GENERALE DEL POTENZIALE RISCHIO CHE L'IA POSSA PORTARE ALLA DISCRIMINAZIONE

Un sondaggio Eurobarometro che includeva domande sull'IA ha chiesto agli intervistati quali sono gli aspetti che li preoccupano maggiormente quando si tratta dell'uso dell'IA, tra cui la discriminazione nel processo decisionale, la responsabilità poco chiara e il fatto che non c'è nessuno con cui lamentarsi.

Solo il 40% circa dei cittadini dell'UE ha dichiarato di temere che l'uso dell'IA possa portare a discriminazioni in termini di età, genere, razza o nazionalità, ad esempio nelle decisioni in materia di assunzioni, affidabilità creditizia, ecc.

I risultati variano da paese a paese. Proporzioni più elevate di persone sono preoccupate per la discriminazione nei Paesi Bassi (58%), Lussemburgo (48%) e Svezia (47%). Proporzioni inferiori hanno espresso preoccupazione in Estonia (25%), Ungheria (24%) e Lituania (23%).

Tuttavia, da questa domanda non risulta chiaro se le persone non sanno che può verificarsi una discriminazione, o se sono consapevoli che ciò può verificarsi ma non pensano che sia un problema.



Note: Include persone che hanno indicato di essere preoccupate che l'AI possa portare a una discriminazione tra tre possibili questioni, o tutte e tre le questioni.

Fonte: calcoli FRA basati sulla Commissione europea (2019), Eurobarometro, 92.3

GENERAZIONE Z

Com'è noto, la Generazione Z, che comprende le classi 1995 - 2010, è la prima generazione nata e cresciuta in un contesto completamente digitale; per questo è particolarmente interessante interrogarsi sul livello di diffusione dell'intelligenza artificiale in questo gruppo e sulla consapevolezza rispetto ai potenziali *bias*. [L'indagine](#) ha confermato un uso molto diffuso dell'AI e una polarizzazione netta sulla consapevolezza.

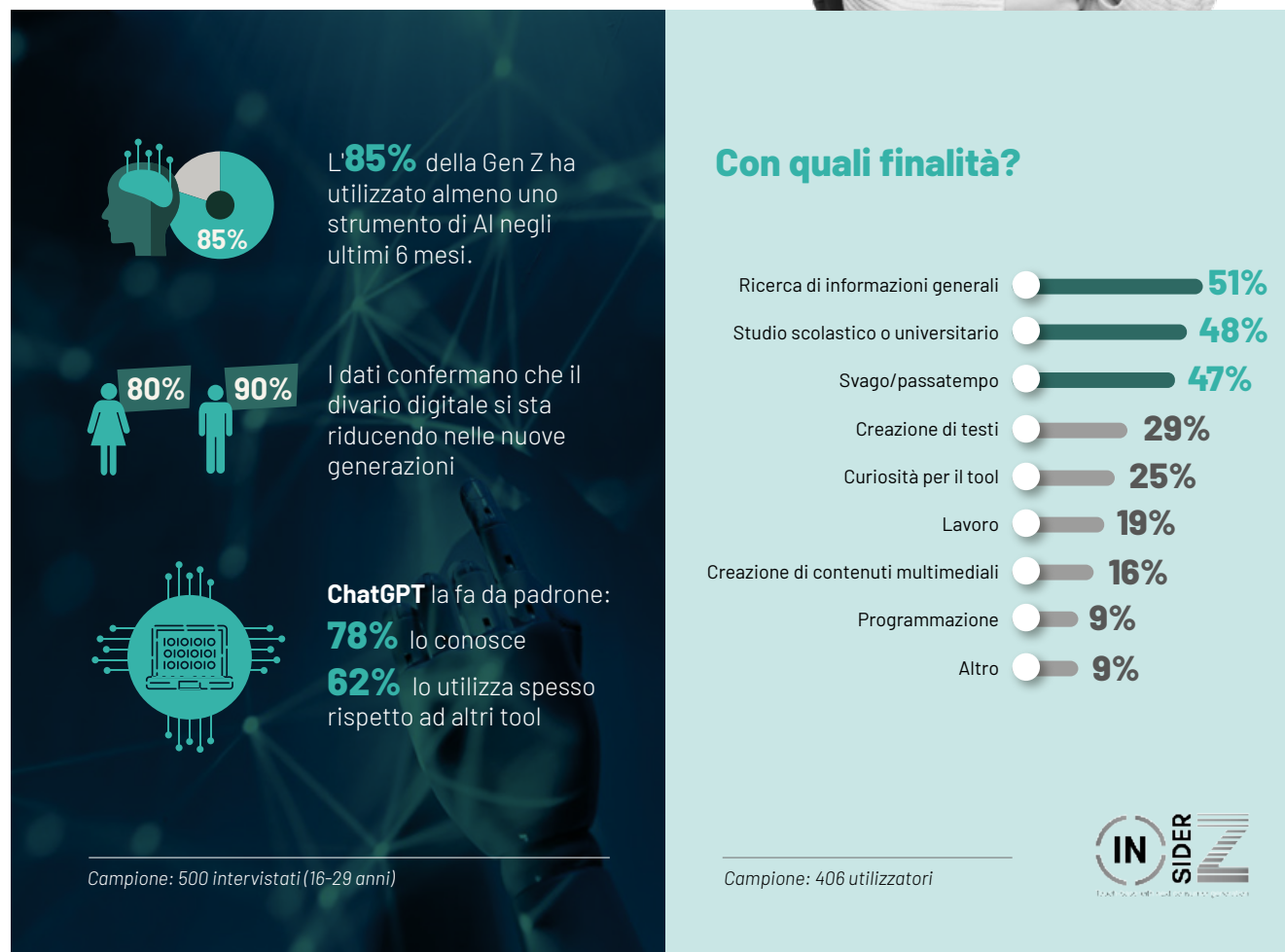
Come si evince dalle infografiche, ChatGBT è l'AI più conosciuta e più usata, seguita da Google e Bing. Strumenti più specifici come MidJourney sono invece ignoti a tutte le fasce d'età. La maggioranza delle persone rispondenti le ha usate negli ultimi 6 mesi, anche se le ragazze sono un po' meno rappresentate dei ragazzi.

Questo dato è molto interessante, dal momento che in altre fasce di età della popolazione il divario digitale di genere è molto più ampio.

Le AI vengono principalmente utilizzate per studiare (fascia 16-21), per lavorare (25-29) e per cercare informazioni; l'utilizzo per svago e curiosità (più maschi che femmine) è altrettanto diffuso.

Più ridotto è il numero di chi utilizza l'AI per attività specifiche (ad esempio per creare testi, immagini, ecc.), il che induce a pensare che siamo ancora in una fase di familiarizzazione e curiosità verso questa tecnologia.

Di seguito i risultati della survey realizzata espressamente per questo report nel 2024 da Fondazione Giacomo Brodolini e Tele Performance Knowledge Services e il loro Osservatorio InsiderZ su un campione di giovani della GEN Z.

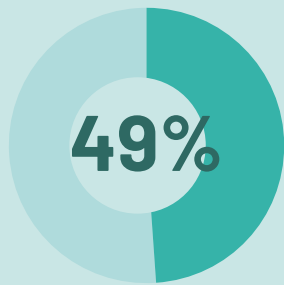


Tuttavia, a un utilizzo molto diffuso non corrisponde un'altrettanta diffusa riflessione sui possibili *bias* discriminatori della nuova tecnologia.

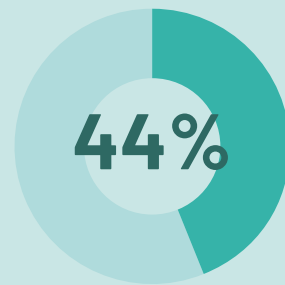
Oltre la metà delle persone di tutte le fasce che hanno risposto al sondaggio non si è mai interrogata sul fatto che questi strumenti possano produrre contenuti condizionati da *bias* discriminatori.

Tra chi ritiene che ci possano essere distorsioni, le aree maggiormente indicate come potenzialmente soggette a *bias* sono quelle del "genere" e dell'"orientamento politico".

Percezione della presenza di bias nell'AI



Totale intervistati



Utilizzatori strumenti AI

Non si è mai posto il dubbio su eventuali bias nell'AI



Possibili distorsioni

38%

Genere

32%

Età, colore della pelle, orientamento sessuale e rappresentazione dei corpi

22%

Neurodivergenze e neurotipicità

20%

Diverse abilità fisiche

36%

Orientamento politico

23%

Religione

Campione: 186 intervistati indicano la presenza di distorsioni

POLITICHE E PRATICHE IN ITALIA E IN EUROPA

LIMITARE IL POTENZIALE EFFETTO DISCRIMINATORIO DELL'AI

La discriminazione è un argomento cruciale quando si parla di utilizzo dell'intelligenza artificiale, **perché lo scopo stesso degli algoritmi di apprendimento automatico è categorizzare, classificare e separare**. Differenziare non è di per sé un'azione negativa; ad esempio, quando si decide di concedere un prestito, la storia creditizia può essere utilizzata per fare una differenziazione fra gli individui, ma non sulla base di caratteristiche sensibili come il sesso o la religione. Tuttavia, la storia creditizia potrebbe essere sistematicamente diversa per uomini e donne a causa delle differenze nei guadagni e nelle storie lavorative. L'utilizzo degli algoritmi potrebbe esasperare queste disuguaglianze, rendendo sempre più difficile ridurle.

Agire sugli algoritmi **potrebbe contribuire a ridurre i pregiudizi e gli stereotipi**, soprattutto se si riuscisse a formulare un algoritmo di complessità tale da **tener conto di tutti gli elementi di compensazione** necessari per contemplare

tutte le variabili relative a diversità e inclusione. Si tratta di un processo lungo, e i primi tentativi in atto non sono del tutto convincenti.

Le persone esperte dicono che siamo ancora in una fase fluida, in cui le potenzialità non sono del tutto realizzate e i processi di *machine learning* non sono completamente comprensibili.

Nel frattempo, però, occorre muoversi su almeno tre fronti:

- interrogarsi, a livello globale, su quali siano i **limiti di azione dell'AI**, non tanto in una prospettiva tecnologica ma politica ed etica;
- rendere **consapevoli cittadini e cittadine del funzionamento di questa nuova tecnologia**, che, come tutte le tecnologie (dall'aratro al PC passando per la macchina a vapore) comporta potenzialità e rischi e richiede conoscenze per un utilizzo consapevole e competente;
- costruire le condizioni per un'**integrazione** tra persone e tecnologia (l'AI) che favorisca il **superamento delle disuguaglianze** invece che contribuire ad accentuarle.



In un [articolo](#) per la rivista ingenerere.it, pubblicata dalla Fondazione Giacomo Brodolini, Ivana Bartoletti, una delle principali esperte di AI e privacy in Europa, ci ricorda che su questi temi serve **una discussione aperta**. L'intelligenza artificiale, a suo avviso, non è affatto neutrale come ci si può aspettare, ed è generalmen-

te declinata al maschile, più che al femminile. E lo è non solo perché gli informatici e i programmatori che la generano sono, in grande maggioranza, maschi; ma anche perché, in sostanza, sta ridisegnando il pianeta a partire dal modello economico, sociale e simbolico dominante, quello che continua a favorire gli uomi-

ni. È fondamentale quindi muoversi verso una **sempre maggiore consapevolezza della dimensione non solo tecnologica dell'AI, e verso una regolamentazione** delle informazioni e dei dati utilizzati.

DONNE CHE PROPONGONO SOLUZIONI INCLUSIVE PER L'INTELLIGENZA ARTIFICIALE

La rivista Wired ha recentemente pubblicato un [articolo](#) dedicato ad alcune delle pioniere che si stanno occupando di progetti volti a risolvere alcune delle questioni aperte dell'AI; altre se ne possono trovare nella lista pubblicata ogni anno da [Women AI Ethics](#).

Tra loro, **RUMMAN CHOWDHURY**, già a capo della divisione di **ricerca etica sull'intelligenza artificiale di Twitter**, ha co-fondato con Jutta Williams [Humane Intelligence](#), un'organizzazione no-profit che offre varie opzioni di testing su larga scala per la cosiddetta intelligenza artificiale generativa, coinvolgendo utenti con profili diversificati per stanare il possibile impatto discriminatorio delle nuove tecnologie.

SARAH BIRD si occupa in Microsoft di evitare che l'intelligenza artificiale generativa che l'azienda sta integrando nelle sue applicazioni per l'ufficio produca discriminazioni; il suo

team lavora per contenere le derive discriminatorie a cui sono esposti i chatbot che apprendono direttamente dai contenuti online.

YEJIN CHOI, docente presso la School of Computer Science & Engineering dell'Università di Washington, sta lavorando allo sviluppo di [Delphi](#), un **modello open source progettato per insegnare alle macchine il senso morale, ovvero ad avere un senso di ciò che giusto e sbagliato**. Il suo obiettivo è creare sistemi con capacità pari a quelle di OpenAI e Google, ma che non richiedano enormi risorse, che possano essere applicati con maggiore facilità e aiutino le macchine a distinguere in termini di senso etico tra due frasi che a un chatbot potrebbero apparire simili, ad esempio "aiutare un'amica" e "aiutare un'amica a diffondere fake news".

INIOLUWA DEBORAH RAJI lavora per la [Mozilla Foundation](#), dove sviluppa strumenti open source per analizzare i sistemi di intelligenza artificiale, compresi i modelli linguistici di grandi dimensioni, e individuarne difetti come pregiudizi e imprecisioni; sostiene che questi strumenti possano **aiutare le comunità danneggiate dall'AI a contestare le tesi portate avanti dalle potenti aziende tecnologiche**.

DANIELA AMODEI è tra le fondatrici di [Anthropic](#), una società che sta adottando un approccio diverso alla sicurezza dell'AI attraverso la progettazione di un chatbot, Claude, dotato di una "costituzione" che guida il suo comportamento, basata su **principi tratti da fonti come la Dichiarazione universale dei diritti umani delle Nazioni Unite**. L'obiettivo è ridurre i comportamenti scorretti, che forse aiuterà a limitare i sistemi AI del futuro, che saranno ancora più potenti.

LA STRATEGIA EUROPEA PER L'INTELLIGENZA ARTIFICIALE E I DIRITTI FONDAMENTALI

Un approccio alla regolamentazione dell'AI è quello espresso nel [Libro Bianco dell'Unione europea sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia](#).

Il Libro Bianco delinea i principi fondamentali di un futuro quadro normativo dell'Unione europea per l'intelligenza artificiale, ed evidenzia come quest'ultimo si fondi **sui valori fondamentali dell'UE, compreso il rispetto dei diritti umani**.

L'obiettivo è "promuovere l'adozione dell'intelligenza artificiale e **affrontare i rischi per i diritti fondamentali associati a determinati usi di questa nuova tecnologia**". Si riconosce che "l'uso dell'AI può incidere sui valori su cui si fonda l'UE e portare a violazioni dei diritti fondamentali, sia a causa di difetti nella progettazione complessiva dei sistemi di AI, sia a causa dell'uso dei dati senza correggere eventuali pregiudizi".

L'obbligo di rispettare il principio di non discriminazione è sancito dai trattati europei e dalla Carta dei diritti fondamentali.

L'automazione e l'uso dell'intelligenza artificiale possono aumentare notevolmente l'efficienza dei servizi e ampliare i compiti che gli esseri umani non sarebbero in grado di svolgere. È necessario tuttavia garantire che i servizi e le decisioni basate sull'intelligenza artificiale non siano discriminatorie. Il documento offre una panoramica dei diritti fondamentali sui quali l'AI può avere un impatto.

DIRITTI FONDAMENTALI A RISCHIO CON UN USO NON DISCIPLINATO DELL'AI

Se il diritto alla protezione **dei dati personali e il diritto alla privacy** entrano in gioco in quasi tutte le applicazioni dell'AI, in alcuni ambiti è importante valutare l'impatto di altri diritti specifici. Ad esempio:

- il diritto alla protezione sociale, quando si lavora con le prestazioni sociali;
- il diritto alla libertà di espressione e informazione, quando si utilizza l'intelligenza artificiale per supportare la moderazione dei contenuti online;
- il diritto di riunione e di associazione, quando si considera l'uso della tecnologia di riconoscimento facciale negli spazi pubblici;
- il diritto all'istruzione, quando si utilizza l'AI nel settore educativo;
- il diritto di asilo, quando si utilizza l'intelligenza artificiale per supportare la gestione della migrazione;
- il diritto alla contrattazione e all'azione collettiva, quando si utilizza l'intelligenza artificiale nella cosiddetta *gig economy*;
- il diritto a condizioni di lavoro eque e giuste quando si utilizza l'intelligenza artificiale sul posto di lavoro;
- il diritto di accesso all'assistenza sanitaria preventiva, quando si utilizza l'AI nei servizi sanitari;
- il diritto alla presunzione di innocenza e il diritto alla difesa, quando si utilizza l'AI nel settore della giustizia.

Attualmente, l'applicazione della legislazione UE sulla protezione dei dati dipende prevalentemente dal trattamento dei dati personali, sebbene **alcune applicazioni basate sull'intelligenza artificiale non utilizzino dati personali** (ad esempio, i dati sul traffico), mentre altre utilizzano dati anonimizzati. In questi casi, le leggi sulla protezione dei dati non si applicano, o la loro applicabilità non è del tutto chiara. Il confine tra dati personali e non personali è sfumato, perché esiste il rischio che i dati anonimizzati possano essere "re-identificati" - vale a dire, che l'anonimizzazione possa essere annullata.

In un recente discorso tenuto nell'ambito della conferenza [Protecting Human Rights in the Digital Age](#), Michael O'Flaherty, l'allora Direttore dell'[Agenzia Europea dei Diritti Fondamentali](#), ha elencato gli elementi che dovrebbero caratterizzare un sistema giuridico europeo di regolamentazione dell'AI che rispetti i diritti umani:

- contemplare **leggi che siano esaustive**, ovvero una regolamentazione senza scappatoie, che parta da una definizione ampia di intelligenza artificiale e si applichi allo stesso modo al settore privato e a quello pubblico. Esiste infatti il rischio di definire in modo restrittivo l'intelligenza artificiale ed escludere elementi come i database che si trovano fuori dai confini e quindi dalla le-

gislazione europea, o anche di considerare soltanto alcuni diritti fondamentali, come ad esempio il diritto alla privacy, tralasciandone altri, come il diritto alla non discriminazione e il diritto di accesso alla giustizia, il diritto alla dignità e la protezione dei consumatori.

- Prevedere **test di compatibilità** con i diritti umani per le applicazioni ad alto rischio. È fondamentale che, laddove un'applicazione presenta un rischio elevato per il benessere umano, venga testata nel contesto d'uso, in modo da poter comprendere quale sia il rischio e gestirlo. È importante che i test vengano ripetuti per contenere il fenomeno dei cicli di feedback e della crescita esponenziale degli effetti non voluti.
- Prevedere una **forte supervisione**, prestare attenzione a garantire che i sistemi in atto per supervisionare la regolamentazione siano adeguati al compito, che dispongano di competenze e risorse. Servono specialisti e specialiste non solo di privacy, ma di tutti i diritti umani, che lavorino a stretto contatto con chi progetta l'AI.
- Prevedere **un rimedio per ogni violazione**, in accordo con i principi fondamentali dei diritti umani: sia nelle normative specifiche per l'intelligenza artificiale sia in una legislazione separata, che è il modello dell'UE in questo momento, è necessario garanti-

re che esista un percorso verso un rimedio per ciascuna persona la cui dignità umana è stata violata da un'applicazione della tecnologia.

- **Garantire la trasparenza:** per un controllo efficace e un adeguato monitoraggio della tecnologia, è fondamentale che ci sia trasparenza riguardo ai suoi contenuti. Questo deve avvenire nonostante le resistenze sia da parte di chi afferma che l'estrema complessità degli algoritmi rende difficile la trasparenza, e sia da parte di chi invoca la sicurezza nazionale. In entrambi i casi, rendere trasparenti i processi garantirebbe il rispetto dei diritti senza necessariamente compromettere la riservatezza sui contenuti.
- **Mantenere un dialogo continuo con le diverse persone portatrici di interessi:** nella progettazione dei regolamenti, nella loro introduzione, applicazione e futura modifica, tutto deve avvenire sulla base di un dialogo ricco e vivo tra tutte le parti interessate.
- **Superare l'assunto secondo cui la protezione dei diritti umani costituirebbe un freno allo sviluppo dell'innovazione**, che si traduce in un rischio per la competitività. È possibile garantire attenzione ai diritti umani pur rispettando profondamente il bisogno di innovazione nelle nostre società, utilizzando un [modello piramidale del rischio](#), come sta facendo l'Europa. Questa pirami-

de presenta le applicazioni più rischiose nella fascia più alta, seguite da un'ampia fascia di applicazioni "ad alto rischio", per le quali si propone una regolamentazione piuttosto severa. Infine, alla base della piramide c'è un'ampia fascia in cui si trovano tutte le applicazioni "benigne" dell'AI, in cui nessuno può incorrere in grandi rischi e che sarebbe soggetta a una supervisione minima sui diritti umani.

IL MANIFESTO PER LA SOSTENIBILITÀ DIGITALE

Pur interessando i vari ambiti di Competenze, Ricerca ed Applicazioni, il [Programma Strategico Intelligenza Artificiale 2022-2024](#) per l'Italia non sembra prevedere un focus specifico sulla lotta alle discriminazioni e della promozione di un'intelligenza artificiale più inclusiva, al netto di un'attenzione al superamento del divario di genere nei percorsi formativi.

Nel settembre del 2023, la Fondazione per la Sostenibilità Digitale ha presentato un [Manifesto per la sostenibilità digitale dell'intelligenza artificiale](#), che riprende i temi affrontati in ambito europeo e propone una serie di elementi di attenzione per lo sviluppo dell'intelligenza artificiale in Italia, collegandoli all'esigenza di raggiungere la **sostenibilità energetica e ambientale**.

CARATTERISTICHE DI UN'AI SOSTENIBILE

Per garantire lo sviluppo sostenibile dell'intelligenza artificiale, è necessario che quest'ultima rispetti alcune caratteristiche intrinseche, che devono regolarne l'evoluzione e che sono concepite per minimizzare i rischi della sua applicazione massimizzando le opportunità dischiuse dall'utilizzo di questa tecnologia:

- ✓ rispetto dei diritti fondamentali;
- ✓ privacy;
- ✓ trasparenza;
- ✓ non discriminatorietà;
- ✓ sicurezza;
- ✓ interoperabilità – ovvero deve basarsi su standard e protocolli aperti, in grado di garantire in maniera ottimale lo scambio e il riutilizzo delle informazioni;
- ✓ portabilità, ovvero l'utente deve avere la possibilità di esportare i propri dati (non soltanto quelli personali) in un formato strutturato e trasferirli, anche in maniera automatizzata, da un sistema all'altro;
- ✓ accessibilità, ovvero garantire, fin dalla fase di progettazione e in quelle successive, l'accesso alle persone con disabilità, sulla base del principio di uguaglianza e "senza lasciare indietro nessuno";
- ✓ revoca, ovvero deve essere garantita la possibilità di un'efficace supervisione umana, che possa incidere laddove necessario su processi e azioni governati o eseguiti dall'AI;
- ✓ riconoscibilità: l'utenza deve essere messa in condizione, in maniera semplice e intuitiva, di sapere che sta interagendo con un sistema di AI;
- ✓ proporzionalità del rischio: nello sviluppo dell'AI deve esistere un rapporto di proporzionalità tra i modelli di implementazione, le dinamiche di utilizzo e i processi di regolamentazione, in relazione alla portata dei rischi che possono essere generati dalla sua adozione;
- ✓ efficienza energetica: i sistemi di AI devono essere progettati tenendo in considerazione l'impatto ambientale generato tanto nelle fasi di addestramento che di esercizio e utilizzo da parte dell'utenza.

RIEPILOGO DELLE PRINCIPALI POLITICHE PREVISTE

Talenti e Competenze	Ricerca		Applicazioni	
			Per le aziende	Per la PA
A.1 Rafforzare il programma Nazionale di Dottorato Aumentare il numero di dottori di ricerca	B.1 Rafforzare l'ecosistema italiano di ricerca sull'IA Creare un'architettura di ricerca su base hub & spoke con competenze territoriali	B.5 Promuovere campioni nazionali IA multidisciplinari Lanciare sfide sui temi specifici con concorrenti valutati sulla base di risultati misurati	D.1 Fare dell'AI un pilastro a supporto della Transizione 4.0 delle imprese Introdurre crediti d'imposta o voucher per l'assunzione di profili STEM; aggiornamento dell'elenco delle spese software e hardware ammissibili agli incentivi transizione 4.0	E.1 Creare interoperabilità e dati aperti per favorire la creazione di modelli di IA Creare interoperabilità tra le banche dei dati della PA e mantenere aggiornate le linee guida per Open Data riutilizzabili per modelli di IA con dataset estesi e annotati
A.2 Attrarre e trattenere ricercatori Attrarre giovani ricercatori beneficiari di borse di ricerca internazionali di altro profilo come l'ERC	B.2 Lanciare la piattaforma italiana di dati e software per la ricerca sull'IA Creare una connessione strutturale di piattaforme nuove ed esistenti, dati e infrastrutture informatiche dedicate all'IA, con librerie open-source	B.6 Lanciare bandi di ricerca-innovazione IA per collaborazioni pubblico-private Promuovere progetti su settori prioritari ma con proposte di libera iniziativa volte a trasferire competenze dalla ricerca alle industrie	D.2 Sostenere la crescita di spin-off innovativi e start-up Promuovere la collaborazione all'interno degli ecosistemi delle start-up: offrire appalti pubblici alle start-up per l'acquisto di beni e servizi	E.2 Rafforzare le soluzioni IA nella PA e nell'ecosistema GovTech in Italia Introdurre bandi periodici per identificare e supportare le start-up con potenziali soluzioni basate sull'IA per efficientare la PA e migliorarne i servizi
A.3 Rafforzare le competenze di IA nella Pubblica Amministrazione Attivare tre cicli di nuovi corsi di dottorato specificatamente progettati per le esigenze generali della PA	B.3 Creare cattedre italiane di ricerca sull'IA Allocare fondi specifici per un unico Principal Investigator (PI), già iscritto ad università o centri di ricerca nazionali, per favorire la collaborazione con industrie ed enti pubblici	C.1 Finanziare ricerca e applicazioni dell'IA creativa Finanziare progetti che integrano la ricerca accademica nel campo di frontiera dell'IA creativa assieme alle sue applicazioni industriali	D.3 Promuovere il go-to-market delle tecnologie IA Promuovere Sperimentazione Italia, uno strumento che consente sperimentazioni attraverso un'esenzione temporanea dalla normativa vigente	E.3 Creare un dataset comune di lingua italiana per lo sviluppo dell'IA Creare una risorsa linguistica aperta e condivisa - una raccolta strutturata di dati digitali da documenti in italiano, disponibili a tutti gratuitamente
A.4 Promuovere corsi e carriere in materie STEM Integrare attività, metodologie e contenuti finalizzati allo sviluppo delle materie STEM nei curricula di tutti i cicli scolastici	B.4 Creare iniziative IA-PRIN per ricerca fondamentale Promuovere bandi dedicati alla ricerca fondamentale sull'IA e sull'IA affidabile	C.2 Promuovere progetti bilaterali per incentivare il rientro in Italia di professionisti Lanciare bandi per progetti incentrati sui temi specifici definiti da priorità italiane cofinanziati da un altro Paese con rientro in Italia di almeno un ricercatore	D.4 Supportare le imprese nella certificazione dei prodotti IA Definire un sistema di governance nazionale a supporto della certificazione dei prodotti di IA che si affacciano sul mercato in ambiti con profili di rischio elevato	E.4 Creare banche dati e analisi basate su IA/NLP per feedback/miglioramento dei servizi Creare dataset annotati e anonimizzati interazioni cittadini-PA per supportare lo sviluppo/integrazione dei fornitori di IA nello sviluppo di servizi PA innovativi
A.5 Espandere l'AI negli ITS ("Istituti Tecnici Superiori") Espandere i corsi di programmazione e includere corsi e stage di AI applicata in tutti i curricula ITS			D.5 Promuovere campagne di informazione sull'IA per le imprese Organizzare azioni di comunicazione e sensibilizzazione sull'IA. Le campagne includeranno la diffusione del Programma strategico nazionale per l'IA agli imprenditori	E.5 Creare banca dati IA/Computer Vision per il miglioramento dei servizi nella PA Creare un dataset annotato di grandi dimensioni con immagini satellitari di paesaggi urbani e rurali, incluse immagini catastali digitalizzate
				E.6 Introdurre tecnologie per condivisione e risoluzione casi trasversali a varie autorità Introdurre tecnologie basate sull'IA per automatizzare lo smistamento e la preparazione delle richieste per l'elaborazione

QUADRO NORMATIVO

Il quadro normativo per la disciplina dell'AI è un quadro in continua evoluzione e fortemente condizionato dalla prospettiva europea.

I suoi Pilastri sono il **GDPR**, la **Dichiarazione europea sui diritti e i principi digitali per il decennio digitale** (2023/C 23/ 01), la **Proposta regolamento comunitario sull'Intelligenza Artificiale**.

Vediamo più da vicino questi tre documenti fondamentali.

La GDPR, uno strumento a tutela di cittadini e cittadine europee

Il **regolamento generale sulla protezione dei dati** (*General Data Protection Regulation*, abbreviato in **GDPR**), ufficialmente **regolamento (UE) n. 2016/679**, è un regolamento dell'Unione

europea in materia di trattamento dati personali e privacy adottato nell'aprile del 2016. Con questo regolamento, la Commissione europea si propone di rafforzare la protezione dei dati personali di cittadini e cittadine dell'Unione e delle persone residenti nell'UE, sia all'interno che all'esterno dei confini, restituendo alla cittadinanza il controllo dei propri dati personali, semplificando il contesto normativo che riguarda gli affari internazionali, unificando e rendendo omogenea la normativa privacy dentro l'UE.

Il testo affronta anche il tema dell'esportazione di dati personali al di fuori dell'Unione europea, e obbliga tutti i titolari di trattamento dei dati, anche quelli con sede legale fuori dall'UE, che trattano quelli di persone residenti nell'UE a osservare e adempiere agli obblighi previsti.



FOCUS SUGLI ELEMENTI CHIAVE DELLA GDPR. IL REGOLAMENTO DEFINISCE:

- *Dato personale comune*: informazioni relative a persona fisica identificata o identificabile. Rispetto alle norme precedenti, la novità risiede nel criterio di identificazione: la persona può essere infatti identificata direttamente o indirettamente, con particolare riferimento a un identificativo come il nome e cognome o l'immagine del volto, un numero di matricola, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale; in un quesito, la Commissione europea risponde che "anche varie informazioni che, raccolte insieme, possono portare all'identificazione di una determinata persona costituiscono dati personali".
 - *Profilo*: si delinea il concetto di profilo (di comportamento, di abitudini, di storia), che diventa dato personale pur non contenendo, di per sé, uno specifico dato identificativo di un determinato soggetto; in particolare, parlando di dato personale, occorre distinguere tra identificatori diretti (ad esempio: codice fiscale, per non parlare di impronta digitale o scansione genetica) e identificatori indiretti (ad esempio: indirizzo IP, geolocalizzazione, data nascita e/o luogo di nascita); tra gli identificatori indiretti che però possono portare alla profilazione precisa e quindi all'identificazione di un soggetto fisico ci sono i metadati.
 - *Dati personali e informazioni aggregate*: la definizione di dato personale è relativa, cioè dipende dal contesto; dunque, un dato cessa di essere personale nel momento in cui diventa informazione aggregata: in un gruppo di 10 persone, è molto probabile che la data di nascita sia un identificatore diretto, mentre in un gruppo di 1 milione di persone quasi sicuramente non lo è. Questi dati personali vengono spesso denominati per distinguerli dalle altre tipologie (vedi sotto);
 - *Dati personali particolari o sensibili*: si considerano in questo modo i dati personali come l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche o l'appartenenza sindacale, i dati relativi alla vita sessuale o all'orientamento sessuale della persona, nonché:
 - » *Dati genetici*: ereditati o acquisiti, ottenuti tramite analisi di DNA ed RNA da un campione biologico della persona fisica in questione;
 - » *Dati biometrici*: come l'immagine facciale, grazie ai quali è possibile identificare una e una sola persona fisica;
 - » *Dati sulla salute*: sia fisica che mentale, passata, presente o futura, ma anche informazioni su servizi di assistenza sanitaria, laddove presenti, indipendentemente dalla fonte, ad esempio, un medico.
 - *Dati personali relativi a condanne penali o reati*: il trattamento dei dati personali relativi a reati o condanne penali deve avvenire sotto il controllo dell'autorità pubblica, o se è autorizzato dal diritto dell'Unione o degli stati membri che preveda garanzie appropriate per i diritti e le libertà delle persone interessate.
- I dati personali devono inoltre essere:
- trattati in modo lecito, corretto e trasparente nei confronti della persona interessata, senza rischiare di essere fraintendibile;
 - raccolti per un determinato fine e in maniera esplicita. Non possono dunque essere usati per scopi diversi da quello iniziale, ad eccezione di ricerca scientifica, storica o fini statistici;
 - adeguati, pertinenti e minimizzati al necessario;
 - esatti aggiornati costantemente e rettificati o cancellati se inesatti;
 - conservati limitatamente al tempo necessario al conseguimento delle finalità per le quali sono trattati e non più, ad eccezione di archiviazione nel pubblico interesse, di ricerca scientifica, storica o fini statistici;
 - trattati in maniera adeguata garantendo sicurezza e protezione da trattamenti non autorizzati o illeciti, da perdita, distruzione o danno accidentale attraverso misure tecniche e organizzative adeguate.

LA DICHIARAZIONE EUROPEA SUI DIRITTI E I PRINCIPI DIGITALI PER IL DECENNIO DIGITALE (2023/C 23/ 01)

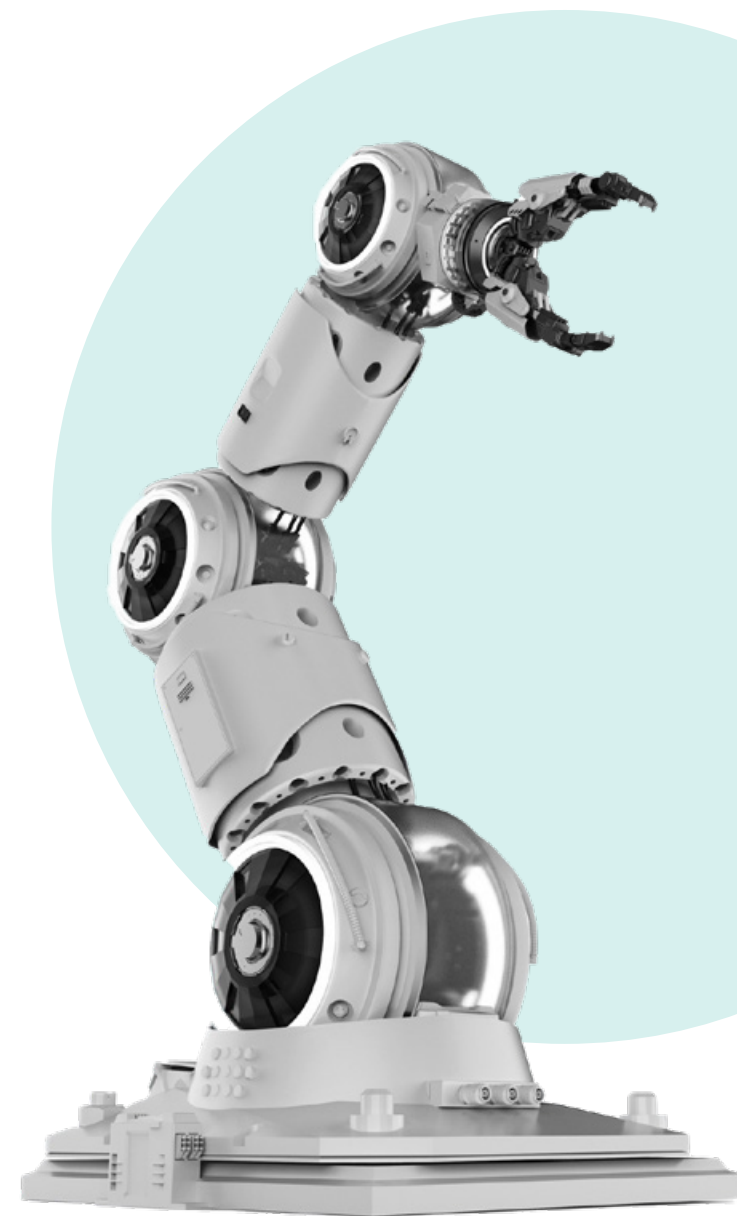
La dichiarazione sui diritti e i principi digitali europei promuove una transizione digitale **antropocentrica all'insegna della sostenibilità** basata interamente **sui valori europei**, in modo tale da garantire a cittadini e cittadine il potere di godere appieno delle opportunità offerte dal decennio digitale. A questi andranno ad aggiungersi nuovi e ulteriori principi e diritti "digitali" che saranno integrati a quelli esistenti e ampiamente garantiti all'interno dell'ordinamento, come la protezione dei dati, l'e-Privacy e la Carta dei diritti fondamentali. Questi diritti prenderanno a modello di riferimento l'esperienza del pilastro europeo dei diritti sociali, e orienteranno sia le stesse istituzioni europee che i singoli stati membri, nel processo di adeguamento alla transizione digitale. I principi si focalizzano su sei temi :

1. porre al centro della trasformazione digitale le persone e i loro diritti.
2. Sostenere solidarietà e inclusione.
3. Garantire la libertà di scelta online.
4. Promuovere la partecipazione allo spazio pubblico digitale.
5. Aumentare la sicurezza, la sicurezza e l'*empowerment* degli individui.
6. Promuovere la sostenibilità del futuro digitale.

IL REGOLAMENTO COMUNITARIO SULL'INTELLIGENZA ARTIFICIALE

Si tratta di una proposta di regolamento su cui è stato raggiunto un accordo dopo un lungo percorso, in quanto prima proposta legislativa di questo tipo al mondo che può fissare uno **standard globale** per la regolamentazione dell'AI in altre giurisdizioni. L'idea principale è quella di regolamentare l'AI sulla base della capacità di quest'ultima di causare danni alla società seguendo un approccio **"basato sul rischio": tanto maggiore è il rischio, quanto più rigorose sono le regole.**

Una volta approvato dal Parlamento europeo e a sei mesi dalla pubblicazione nella gazzetta europea, entreranno in vigore le restrizioni sulle applicazioni AI ad alto rischio "proibite"; ad un anno inizierà la vigilanza sui grandi algoritmi AI, e a due anni entrerà in vigore il marchio CE sugli algoritmi AI.



GLI ELEMENTI CHIAVE

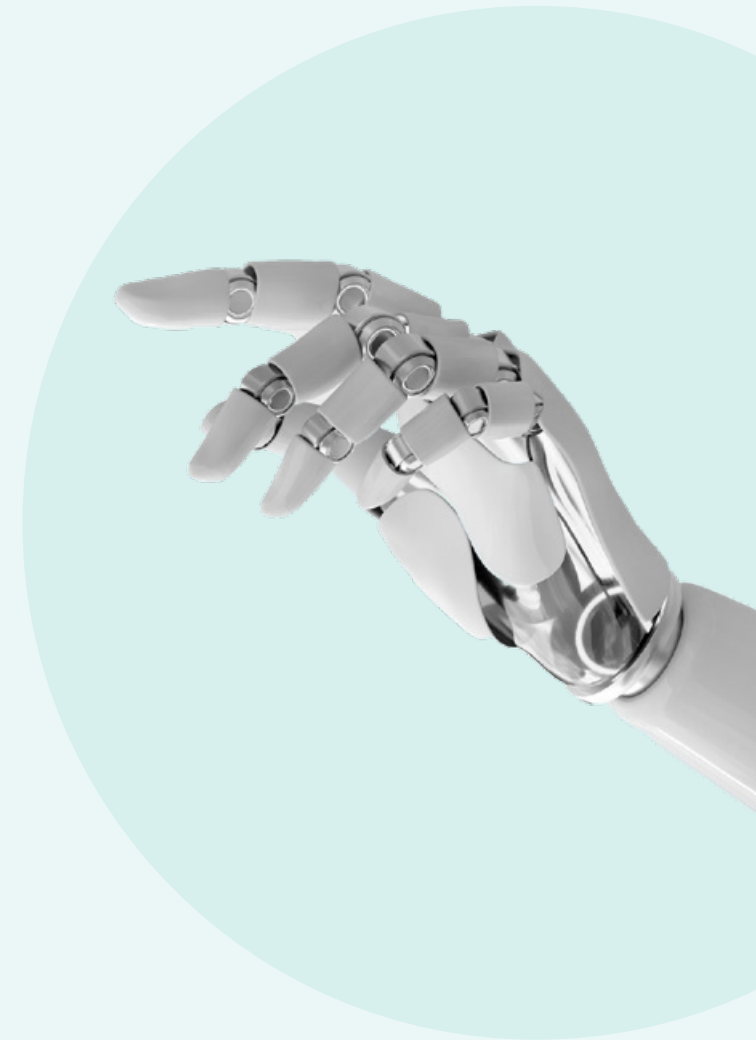
- L'accordo non si applica a settori che non rientrano nell'ambito di applicazione del diritto dell'UE, come la **sicurezza nazionale** e i sistemi utilizzati esclusivamente per scopi **militari** o di **difesa**. Analogamente, l'accordo prevede che il regolamento non si applichi ai sistemi di AI utilizzati solo a scopo di **ricerca e innovazione** o alle persone che utilizzano l'AI per motivi non professionali.
- L'accordo stabilisce un **livello orizzontale di protezione**, compresa una classificazione ad alto rischio, al fine di garantire che non siano inclusi i sistemi di AI che non presentano il rischio di causare gravi violazioni dei diritti fondamentali o altri rischi significativi. I sistemi di intelligenza artificiale che presentano solo un **rischio limitato** sarebbero soggetti a **obblighi di trasparenza** molto leggeri, ad esempio rendere noto che il contenuto è stato generato dall'AI, affinché gli utenti possano prendere decisioni informate in merito all'ulteriore utilizzo.
- Per alcuni usi dell'IA, il rischio è considerato **inaccettabile**; pertanto, tali sistemi saranno vietati dall'UE. L'accordo provvisorio vieta, ad esempio, la **manipolazione comportamentale** cognitiva, lo **scraping** non mirato delle immagini facciali da internet o da filmati di telecamere a circuito chiuso, il **riconoscimento delle emozioni** sul luogo di lavoro e negli istituti di istruzione, il **punteggio sociale**, la **categorizzazione biometrica** per dedurre dati sensibili come l'orientamento sessuale o le convinzioni religiose, e alcuni casi di **polizia predittiva** per le persone.
- L'accordo chiarisce l'**assegnazione delle responsabilità** e i **ruoli dei vari attori** dei sistemi di AI. Chiarisce, inoltre, il rapporto tra le responsabilità a titolo del regolamento sull'intelligenza artificiale e le responsabilità già esistenti in virtù di altri atti legislativi, come la pertinente normativa dell'UE in materia di protezione dei dati o la normativa settoriale.
- Sono previste disposizioni per tenere conto delle situazioni in cui i sistemi di intelligenza artificiale possono essere utilizzati per molti scopi diversi (**AI per finalità generali**) e di quelle in cui la tecnologia di intelligenza artificiale per finalità generali è successivamente integrata in un altro sistema ad alto rischio. L'accordo provvisorio tratta anche i casi specifici dei sistemi di intelligenza artificiale per finalità generali.
- L'accordo prevede che i modelli di base,



ovvero i grandi sistemi in grado di svolgere con competenza un'ampia gamma di compiti distintivi, (la generazione di video, testi, immagini, la conversazione in linguaggio laterale, il calcolo di dati o la generazione di codici informatici) debbano rispettare specifici **obblighi di trasparenza** prima di essere immessi sul mercato. È stato introdotto un regime più rigoroso per i modelli di base **ad alto impatto**

- Un **gruppo scientifico di esperti ed esperte indipendenti** fornirà consulenza all'ufficio per l'intelligenza artificiale in merito ai modelli di AI per finalità generali, contribuendo allo sviluppo di metodologie per valutare le capacità dei modelli di base, fornendo consulenza sulla designazione e l'emergere di modelli di base ad alto impatto e monitorando i possibili rischi materiali di sicurezza connessi ai modelli di base.
- Sono previste **sanzioni pecuniarie** per le violazioni del regolamento sull'intelligenza artificiale, fissate in percentuale del fatturato annuo globale nell'esercizio finanziario precedente della società che ha commesso il reato o, se superiore, in un importo predefinito.

- L'accordo di compromesso chiarisce inoltre che una persona fisica o giuridica può presentare un reclamo alla pertinente **autorità di vigilanza del mercato** riguardo alla non conformità con il regolamento sull'intelligenza artificiale e può aspettarsi che tale reclamo sia trattato in linea con le procedure specifiche di tale autorità.
- L'accordo provvisorio prevede una **valutazione d'impatto sui diritti fondamentali** prima che un sistema di AI ad alto rischio sia immesso sul mercato dai suoi operatori. L'accordo provvisorio prevede inoltre una maggiore **trasparenza** per quanto riguarda l'uso dei sistemi di AI ad alto rischio. In particolare, alcune disposizioni della proposta della Commissione sono state modificate per indicare che alcuni utenti di un sistema di intelligenza artificiale ad alto rischio che sono entità pubbliche saranno anche tenuti a registrarsi nella **banca dati dell'UE** per i sistemi di AI ad alto rischio. Inoltre, nuove disposizioni pongono l'accento sull'obbligo per gli utenti di un **sistema di riconoscimento delle emozioni** di informare le persone fisiche quando sono esposte a tale sistema.



RISORSE PER APPROFONDIRE

BIBLIOGRAFIA

Bartoletti I., *An artificial revolution. On Power, Politics and Ai*, Londra, The Indigo Press, 2020.

Chen Q., Lee K.F., *AI 2041. Scenari dal futuro dell'intelligenza artificiale*, Roma, Luiss University Press, 2023.

Numerico T., *Big data e algoritmi*, Roma, Carocci Editore, 2021.

Rossi F., *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, Milano, Feltrinelli, 2019.

Umoja Noble S., *Algorithms of oppression. How search engines reinforce racism*, New York, New York University Press, 2018.

SITOGRAFIA

AA.VV., *Ada Lovelace Institute* su www.adalovelaceinstitute.org, s.d.

AA.VV., *Agenda digitale* su www.agendadigitale.eu, s.d.

AA.VV., *Artificial intelligence: l'era dell'implementazione* su www.osservatori.net, 2023.

AA.VV., *Artificial Intelligence Platform work and gender equality* su www.eige.europa.eu, 2022.

AA.VV., *Can machines learn morality? The Delphi experiment* su www.arxiv.org, 2022.

AA.VV., *Ethics guidelines for trustworthy AI* su www.digital-strategy.ec.europa.eu, 2019.

AA.VV., *European enterprise survey on the use of technologies based on artificial intelligence - Final report* su www.op.europa.eu, 2020.

AA.VV., *Getting the future right - Artificial intelligence and fundamental rights* su www.fra.europa.eu, 2020.

AA.VV., *Glossario* su www.ingenere.it, s.d.

AA.VV., *Il testo del regolamento* su www.garanteprivacy.it, s.d.

AA.VV., *Manifesto per la sostenibilità digitale dell'intelligenza artificiale* su www.camera.it, 2023.

AA.VV., *Programma strategico Intelligenza Artificiale 2022-2024* su www.docs.italia.it, s.d.

AA.VV., *Regolamento sull'intelligenza artificiale: il Consiglio e il Parlamento raggiungono un accordo sulle prime regole per l'IA al mondo* su www.consilium.europa.eu, 2023.

De Micheli B., *Oppresse dagli algoritmi* su www.inegnere.it, 2023.

Forlè R., *Ivana Bartoletti, il futuro dei dati* su www.ingegnere.it, 2023.

Gil D., *The future of AI is too important to be decided behind closed doors. There is a better way* su www.fortune.com, 2023.

Knight W., Johnson K., Meaker M., *Le donne che stanno cambiando l'intelligenza artificiale* su www.wired.it, 2023.

Kusner M.J., Loftus J.R., *The long road to fairer algorithms* su www.nature.com, 2020.

La Trofa F., *L'intelligenza artificiale: cos'è, a cosa serve, esempi di applicazione* su www.tech4future.info, 2022.

Nardocci C., *Intelligenza artificiale e discriminazioni*, Convegno annuale dell'associazione Gruppo di Pisa "Il diritto costituzionale e le sfide dell'innovazione tecnologica" su www.gruppodipisa.it, 2021.

Unesco, International Research Centre on Artificial Intelligence, *Challenging systematic prejudices: an investigation into bias against women and girls in large language models* su www.unesdoc.unesco.org, 2024.



ALCUNI NOSTRI PROGETTI

MASTER IN GENDER EQUALITY DIVERSITY E INCLUSION

Le organizzazioni in cui viviamo e lavoriamo sono sempre meno uniformi per genere, provenienza etnico-culturale, orientamento sessuale, fede religiosa, presenza di persone appartenenti a diverse fasce generazionali o con disabilità fisiche e/o neurodivergenze. In una società multietnica e complessa la diversità non è solo un elemento con cui è necessario sapersi confrontare ma è anche una risorsa che permette di progettare soluzioni creative ed innovative, che tengano conto dei diversi punti di vista. Per poter definire politiche di intervento e strumenti efficaci, nei diversi ambiti sia pubblici che privati, diventa fondamentale avere consapevolezza delle normative anti-discriminatorie, delle dinamiche organizzative e degli strumenti più adatti per definire ed attuare un approccio di diversity management, ovvero di valorizzazione sistematica delle differenze.

Nel programma del Master, giunto alla sua Sedicesima edizione viene dato ampio spazio alle tematiche di genere. Le strutture organizzative vengono, infatti, influenzate a tutti i livelli dalla costruzione sociale dei ruoli di genere, in maniera non sempre esplicita. In questo campo ci sono approcci sperimentati e pratiche di successo che verranno studiati e analizzati e forniranno il punto di partenza per un discorso più ampio sulla valorizzazione delle differenze. Il percorso formativo si articola in otto moduli tematici che si susseguono in un percorso logico, con l'obiettivo di fornire ai/alle partecipanti gli strumenti per conoscere, progettare, valutare e comunicare le politiche di diversity management e gender mainstreaming. I campi privilegiati di analisi e progettazione sono le dinamiche organizzative, nelle organizzazioni pubbliche e private, e le politiche di welfare, nella loro accezione più ampia.

mastergedm.it



TARGET (2018-2022)

TARGET contribuisce all'avanzamento della parità di genere nella ricerca e nell'innovazione attraverso il sostegno di una politica di parità di genere partecipativa e riflessiva in sette istituzioni dell'area mediterranea inclusi enti di ricerca, enti che finanziano ricerca e un network di università.

L'approccio di TARGET va oltre l'adozione formale di una politica di pari opportunità: attraverso un processo interattivo e riflessivo che porta a una parità a livello istituzionale e il rafforzamento di una comunità di pratica interna a ogni istituzione. I cambiamenti infatti avvengono qualora aumenti la capacità istituzionale di identificare, riflettere su e contrastare le discriminazioni di genere. Il punto di partenza e pilastro dell'approccio di TARGET è quello di produrre per ogni istituzione coinvolta un piano per la parità pensato specificamente per il contesto. I piani per la parità verranno progettati, implementati, monitorati e valutati nell'arco del progetto.

GRASE. GENDER AND RACE STEREOTYPES ERADICATION IN LABOUR MARKET ACCESS (2021-2022)

GRASE - Eradicazione degli Stereotipi di Genere e Razza nell'accesso al mercato del lavoro è un progetto biennale finanziato dall'UE nell'ambito del programma europeo REC

GRASE mira a ridurre le barriere nel sistema dei servizi di consulenza e orientamento professionale per facilitare l'accesso delle donne con background migratorio al mercato del lavoro e a posizioni e settori appropriati alle loro competenze.

L'approccio di GRASE mira ad attivare collaborazioni sinergiche per affrontare queste sfide in una prospettiva intersezionale.



CREDITS:

La foto di Giacomo Brodolini a pag. 3
è rilasciata sotto licenza [Creative Commons Attribution 3.0 Italy License](#)
Fonte: [Wikipedia](#) via [senato.it](#)



The image features a teal gradient background with a circular graphic element in the center. The text is centered within this circle.

***Future of
Workers***