ISSN: 2386-9658 https://doi.org/10.21071/arf.v37i.18691

Intelligenza Artificiale e Fake News: Analisi Linguistica, Gender Bias ed Etica nella comunicazione digitale dei media

(Artificial Intelligence and Fake News: Linguistic Analysis, Gender Bias, and Ethics in Digital Media Communication)

JULIA MARY SCILABRA
https://orcid.org/0009-0008-1314-1212
juliamsc@ucm.es
Universidad Complutense de Madrid https://ror.org/02p0gd045

DARIO RUSSO
https://orcid.org/0000-0003-0705-3028
dariorusso85@uma.es
Universidad de Málaga https://ror.org/036b2ww28

Fecha de recepción: 11 de octubre de 2025
Fecha de aceptación: 10 de febrero de 2026

Riassunto: Lo scenario digitale rappresenta uno spazio privilegiato per la manipolazione e la conseguente diffusione di contenuti falsi o mistificati, spesso amplificando bias di genere e consolidando stereotipi attraverso un linguaggio improprio. Inoltre, l'impiego di strumenti di intelligenza artificiale (IA) per la generazione, moderazione e distribuzione di contenuti sembra non solo riprodurre, ma spesso rafforzare tali bias, contribuendo in tal modo alla costruzione di narrazioni che perpetuano stereotipi e disinformazione.

Adottando un approccio misto, il presente studio combina analisi linguistica e cross-linguistica con prospettive tecnico-teoriche sulla comunicazione digitale ed etica dell'IA, inserendosi nel dibattito contemporaneo sulla capacità delle tecnologie generative di amplificare bias e stereotipi. In tale cornice, la ricerca mira a esplorare come la dimensione quantitativa – relativa alla frequenza e distribuzione dei termini marcati – e quella qualitativa – legata alle dinamiche discorsive e semantiche – possano integrarsi per restituire un quadro più completo dei processi di mistificazione e rappresentazione di genere nei media digitali, partendo da un corpus di 12 articoli che si concentra sui due temi più ricercati in Regno Unito e Stati Uniti, selezionati tramite *Google Trends 2024*. Il corpus comprende articoli del *The New York Times*, *The Washington Pos*t, *The Guardian* e *Daily Mirror*, testate in cui è stato possibile riscontrare un utilizzo

dichiarato di tecnologie di intelligenza artificiale. I risultati rivelano che, pur con differenze fra testate, il linguaggio mediato da algoritmi non è neutro e l'uso di espressioni contrastive o la scelta di forme passive tendono a rafforzare dinamiche discriminatorie.

Parole chiave: IA. Fake news. Gender bias. Terminologia. Giornalismo.

Abstract: Often amplifying gender bias and consolidating stereotypes through inappropriate language, the digital landscape is a privileged space for the manipulation and dissemination of false or misleading content. Moreover, the deployment of Artificial Intelligence (AI) systems in content generation, moderation, and dissemination not only appears to reproduce pre-existing biases but also to intensify them, thereby contributing to the consolidation of narratives that perpetuate stereotypes and facilitate the circulation of disinformation.

Adopting a mixed-method approach, this exploratory study combines corpus-based linguistic and cross-linguistic analysis with technical and theoretical perspectives on digital communication and AI ethics, positioning itself within the contemporary debate on the capacity of generative technologies to amplify bias and stereotypes. Within this framework, the research integrates three analytical dimensions: (1) a quantitative analysis measuring the frequency and normalized distribution of gender-marked and non-inclusive terms; (2) a sentiment analysis to examine how emotional polarity (positive, neutral, negative) interacts with lexical bias and framing; and (3) a qualitative semantic and pragmatic examination of discursive features, including contrastive constructions, voice and agency patterns, and semantic role attribution. The study is based on a corpus of 12 articles focusing on the two most-searched topics in the United Kingdom and the United States, selected using Google Trends 2024. The corpus includes articles from *The New York Times, The Washington Post, The Guardian, and Daily Mirror,* newspapers in which a declared use of Artificial Intelligence technologies was identified. The findings reveal that, despite differences among newspapers, algorithmically mediated language is not neutral, and the use of contrastive expressions or the choice of passive constructions tends to reinforce discriminatory dynamics.

Keywords: AI. Fake news. Gender bias. Terminology. Journalism.

Resumen: El escenario digital constituye un espacio privilegiado para la manipulación y la consiguiente difusión de contenidos falsos o distorsionados, que a menudo amplifican sesgos de género y consolidan estereotipos a través de

un uso inapropiado del lenguaje. Además, el empleo de herramientas de inteligencia artificial (IA) para la generación, moderación y distribución de contenidos no solo parece reproducir, sino con frecuencia reforzar dichos sesgos, contribuyendo así a la construcción de narrativas que perpetúan estereotipos y desinformación.

Adoptando un enfoque mixto, el presente estudio combina análisis lingüístico y translingüístico con perspectivas técnico-teóricas sobre la comunicación digital y la ética de la IA, insertándose en el debate contemporáneo acerca de la capacidad de las tecnologías generativas para amplificar sesgos y estereotipos. En este marco, la investigación busca explorar cómo la dimensión cuantitativa —relativa a la frecuencia y distribución de los términos marcados— y la cualitativa —vinculada a las dinámicas discursivas y semánticas— pueden integrarse para ofrecer una visión más completa de los procesos de mistificación y representación de género en los medios digitales.

El corpus de análisis está compuesto por 12 artículos centrados en los dos temas más buscados en Reino Unido y Estados Unidos, seleccionados a través de *Google Trends 2024*. Dicho corpus incluye artículos de *The New York Times*, *The Washington Post*, *The Guardian* y *Daily Mirror*, periódicos en los que ha sido posible constatar un uso declarado de tecnologías de inteligencia artificial. Los resultados revelan que, aun con diferencias entre las cabeceras, el lenguaje mediado por algoritmos no es neutro: el recurso a expresiones contrastivas o a estructuras pasivas tiende a reforzar dinámicas discriminatorias.

Palabras clave: IA. Noticias falsas. Sesgo de género. Terminología. Periodismo.

1. Introduzione

L'avvento dell'era digitale ha trasformato radicalmente il sistema dell'informazione ed il modo di produrre contenuti, creando da un lato nuove opportunità per la diffusione democratica della conoscenza e nuove possibilità in ambito giornalistico (Esser & Neuberger 2018: 194-197) ma dall'altro sono sopraggiunti nuovi rischi (Parcu 2019: 91-109) circa la manipolazione e la distorsione dell'informazione stessa. In questo contesto, l'incremento delle fake news rappresenta una delle sfide contemporanee più complesse da contrastare, soprattutto se si tiene conto anche delle minacce e delle opportunità legate all'intelligenza artificiale (Allcott & Gentzkow 2017: 211-236). Pertanto, il sistema mediatico ha raggiunto livelli di complessità senza precedenti, in cui la distinzione tra informazione e disinformazione diventa sempre più labile (Vosoughi *et al.* 2018: 1146-1151).

Le ricerche nell'ambito della comunicazione digitale hanno evidenziato come l'ecosistema informativo contemporaneo sia particolarmente vulnerabile di fronte alle *fake news* e che la velocità di propagazione delle notizie false sui social media superi significativamente quella delle informazioni verificate (Tandoc *et al.* 2018: 137-153), così da generare un impatto sistemico sui processi democratici e sulla formazione dell'opinione pubblica (Lazer *et al.* 2018: 1094-1096). In questo scenario, l'intelligenza artificiale assume un ruolo ambivalente offrendo strumenti per il fact-checking (Shu *et al.* 2017: 22-36) ma anche per la creazione di fake news sempre più sofisticate e difficili da individuare (Chesney & Citron 2019: 1753-1820).

Diversi studi hanno evidenziato come la mistificazione dei contenuti e il ricorso a bias di genere vengano perpetrati e rafforzati quando si ricorre a strumenti di intelligenza artificiale per generare, moderare e diffondere contenuti (Minucci *et al.* 2022; Al-Asadi & Tasdemir 2022). Questa dinamica è particolarmente preoccupante poiché gli algoritmi di IA, addestrati su dataset che riflettono inevitabilmente i pregiudizi presenti nella società, tendono a replicare e amplificare questi bias, contribuendo alla costruzione di narrazioni che perpetuano stereotipi e discriminazioni (Nazar & Bustam 2020; Bolukbasi *et al.* 2016).

In questo contesto, la questione del gender bias rappresenta un aspetto cruciale. Molti studi dimostrano come i sistemi di IA possano sistematicamente sottorappresentare o rappresentare in maniera distorta le donne e le minoranze di genere, sia attraverso la selezione di immagini stereotipate sia mediante l'uso di linguaggio che riflette pregiudizi sociali radicati (Marchetti 2019: 107-115; Caliskan *et al.* 2017: 183-186). Questi bias (Lopez 2021) hanno implicazioni profonde per la costruzione dell'identità di genere e per la perpetuazione di disuguaglianze strutturali nella società digitale (Noble 2018).

L'intersezione tra intelligenza artificiale, fake news e bias di genere assume particolare rilevanza quando si presta alla strumentalizzazione politica venendo spesso utilizzata come strumento di polarizzazione sociale e mobilitazione politica (Guess *et al.* 2019; Tucker *et al.* 2018). Inoltre, gli algoritmi dei social media amplificano questi effetti creando delle casse di risonanza che rafforzano pregiudizi esistenti e facilitano la diffusione di narrative false o distorte (Sunstein 2017).

Un caso emblematico di queste dinamiche è rappresentato dalle controversie che hanno coinvolto Imane Khelif, pugile algerina che ha partecipato alle Olimpiadi di Parigi del 2024; la Khelif, affetta da iperandrogenismo, è stata al centro di una massiccia campagna di disinformazione caratterizzata dalla diffusione di *rumors* infondati riguardanti una presunta identità transgender (Brik

2025). Questo episodio illustra perfettamente come la combinazione tra algoritmi di amplificazione dei social media e bias di genere possa generare narrazioni false che si propagano in modo virale, alimentando dibattiti polarizzati basati su informazioni errate (Freelon *et al.* 2020). La vicenda evidenzia inoltre come le fake news legate al gender bias possano essere strumentalizzate politicamente, trasformando questioni mediche e sportive complesse in armi di divisione sociale e propaganda ideologica.

Dal punto di vista linguistico, diversi studi hanno evidenziato come il linguaggio utilizzato nella diffusione di fake news sia intriso di un linguaggio più emotivo e polarizzante rispetto alle notizie autentiche (Pérez-Rosas *et al.* 2018: 3391-3401) e che l'impatto sociale e politico della disinformazione legata al gender bias porti a vere e proprie campagne di disinformazione mirate a delegittimare movimenti sociali e diritti civili (Marwick & Lewis 2017), con l'aggravante dell'utilizzo di strumenti di IA nella moderazione dei contenuti, complice involontaria nel perpetuare discriminazioni di genere (Biju & Gayathri 2023).

Infine, la crescente evoluzione degli strumenti di IA generativa e dalla loro democratizzazione, rende sempre più accessibile la produzione di contenuti falsi convincenti (Floridi *et al.* 2020: 689-707) ed allo stesso tempo l'evoluzione degli algoritmi dei social media e la crescente polarizzazione del dibattito pubblico digitale creano condizioni sempre più favorevoli alla diffusione di disinformazione mirata (Roth & Pickles 2020) che purtroppo trovano terreno fertile molto spesso anche tra gli organi di informazione più accreditati (Nakov & Da San Martino 2021) e testate giornalistiche (Shor *et al.* 2019: 526-550).

Pertanto, attraverso un'analisi comparativa di un corpus di dodici articoli pubblicati da quattro tra le più rilevanti testate giornalistiche del panorama internazionale, il presente studio esplorativo si propone di esaminare i meccanismi attraverso cui la scelta terminologica e discorsiva può influenzare la percezione pubblica, rafforzare pregiudizi di genere oppure, al contrario, promuovere una comunicazione più inclusiva ed eticamente consapevole. In tal senso, l'obiettivo non è soltanto descrittivo, ma anche critico. Lo studio intende promuovere una riflessione etica e terminologica su come parole ed algoritmi si intrecciano nel plasmare le rappresentazioni sociali di genere e su come episodi di disinformazione possono essere amplificati o costruiti intenzionalmente per fini ideologici o propagandistici. Lo studio mira così a stimolare un dibattito più consapevole e informato su queste dinamiche complesse.

2. Materiali e Metodi

Attraverso un'analisi comparativa di un corpus di 12 articoli online, selezionati sulla base dei temi riportati da *Google Trends* come i più ricercati nel 2024 in *UK* e negli *USA*. I due paesi sopracitati, Stati Uniti e Regno Unito, sono infatti accomunati dalle stesse chiavi di ricerca più frequenti nell'anno 2024: il neoeletto presidente degli Stati Uniti, Donald Trump, e Imane Khelif, pugile algerina che ha preso parte alle Olimpiadi di Parigi del 2024. Affetta da iperandrogenismo, nel corso delle Olimpiadi, la figura di Khelif è stata al centro di forti discussioni e speculazioni, soprattutto a seguito di *rumors* riguardanti una possibile identificazione come atleta transgender. L'interesse suscitato da questi due temi rappresenta un elemento significativo per l'analisi in corso, poiché evidenzia la convergenza tra dinamiche politico-sociali e questioni legate alla rappresentazione di genere, riflettendo le principali tendenze di ricerca online e i dibattiti sociali nel contesto statunitense e britannico.

I quotidiani oggetto del presente studio sono stati selezionati sulla base del *Digital News Report 2024* del *Reuters Institute for the Study of Journalism (University of Oxford)*, che fornisce una mappatura aggiornata delle principali testate internazionali e delle relative pratiche editoriali digitali. È necessario precisare che, per quanto a conoscenza degli autori, allo stato attuale della ricerca, non esistono strumenti metodologicamente affidabili e universalmente riconosciuti in grado di determinare con assoluta certezza se un testo (o un articolo giornalistico, in questo caso) sia stato interamente o parzialmente generato da sistemi di intelligenza artificiale (Liang *et al.* 2023; Weber-Wulff *et al.* 2023). I cosiddetti *AI detectors*, infatti, presentano limiti significativi in termini di accuratezza, replicabilità e validità scientifica, rendendo problematica qualsiasi attribuzione automatica e univoca della paternità testuale. Inoltre, strategie semplici, quali parafrasi, strumenti di "umanizzazione" del testo e interventi di editing leggero, permettono di ridurre significativamente i tassi di rilevamento (Liang *et al.* 2023; Weber-Wulff *et al.* 2023; Weidinger *et al.* 2021).

Per tale motivo, la scelta dei quotidiani si è concentrata esclusivamente su quelle testate che dichiarano esplicitamente di utilizzare strumenti di intelligenza artificiale nel processo di produzione, di supporto o di ottimizzazione dei contenuti giornalistici. Il campione di studio si è circoscritto a *The New York Times*, *The Washington Post*, *The Guardian* e *Daily Mirror* (ved. Tabella 1).

Di queste testate, sono stati presi in esame i primi tre articoli pubblicati da ciascuna di esse tra il 26 luglio e l'11 agosto 2024, ossia l'arco temporale delle Olimpiadi di Parigi. I filtri di selezione impostati per individuare gli articoli sono stati: allintitle: "Imane Khelif" site: [nome giornale].

Testata completa	N. articoli	Lunghezza stimata per articolo (parole)	Totale stimato (parole)	Media per articolo (parole)
The Guardian	3	600 – 2.400	4.500 – 5.300	1.500 – 1.750
Daily Mirror	3	800 – 1.400	2.800 – 3.700	930 – 1.230
The New York Times	3	1.700 – 2.300	5.400 – 6.600	1.800 – 2.200
The Washington Post	3	1.500 – 2.200	5.000 – 6.200	1.660 – 2.060
TOTALE COMPLESSIVO	**12**	—	**17.700 – 21.800**	**1.475 – 1.815**

Tabella 1- Ripartizione del numero di parole per articolo.
Fonte: elaborazione propria

La scelta di selezionare i primi tre articoli, in ordine di pubblicazione, per ciascuna delle quattro testate online risponde a una combinazione di criteri di sistematicità, comparabilità e riduzione dei bias di selezione.
In primo luogo, l'adozione dell'ordine di pubblicazione come criterio di inclusione consente di evitare una selezione arbitraria o guidata a posteriori dai contenuti, garantendo che il campione non sia influenzato da valutazioni soggettive di rilevanza, qualità o orientamento editoriale degli articoli. In questo senso, il criterio temporale rappresenta una regola replicabile e trasparente, coerente con approcci di campionamento sistematico ampiamente utilizzati negli studi sui media offline ed online.

In secondo luogo, la selezione di un numero uguale di articoli per ciascuna testata (n = 3) assicura una rappresentazione bilanciata delle fonti, evitando che eventuali differenze nella frequenza di pubblicazione o nella produttività editoriale di una singola testata influenzi in modo sproporzionato il campione complessivo. Questo consente un confronto più equo tra le diverse realtà editoriali incluse nello studio.

Infine, la scelta di limitare il campione a 12 articoli complessivi risponde a esigenze di profondità analitica così da avere un campione contenuto ma strutturalmente omogeneo che permetta un'analisi dettagliata e coerente dei contenuti, pur mantenendo al contempo un livello adeguato di varietà informativa. Nel complesso, la strategia di campionamento adottata mira a massimizzare trasparenza, replicabilità e controllo dei bias, risultando appropriata rispetto agli obiettivi esplorativi e comparativi dello studio.

Uno degli aspetti più delicati della ricerca ha riguardato la definizione e selezione dei termini discriminatori e non inclusivi. A tal fine, sono state prese in considerazione le definizioni fornite dall'*European Institute for Gender Equality*

(EIGE 2016), dalle linee guida dell'*American Psychological Association* (APA 2020) e dai documenti dell'UNESCO (UNESCO 2019), che convergono nell'indicare come discriminatorio il linguaggio che svaluta direttamente individui o gruppi, e come non inclusivo quello che, pur senza insulti espliciti, opera un'esclusione semantica o simbolica.

Inoltre, ai fini di esaustività, nel presente studio è stat considerata una dimensione distinta ma concettualmente correlata, ovvero l'uso di linguaggio mascolinizzante (*masculine-coded language*). Come evidenziato da Caliskan *et al.* (2022) e Gaucher *et al.* (2011), infatti, questo tipo di linguaggio non si manifesta attraverso termini apertamente discriminatori o non inclusivi, bensì attraverso pattern lessicali e verbali associati a concetti di agenzia, controllo e potere, contribuendo in modo implicito a rafforzare l'associazione tra successo, competenza e mascolinità.

Coerentemente con quanto riscontrato nella letteratura esistente sul bias linguistico implicito, che distingue tra forme di esclusione lessicalmente marcate e pattern discorsivi che operano a livello strutturale e valutativo (Stahlberg *et al.* 2007; Blodgett *et al.* 2020), il linguaggio mascolinizzante è stato trattato come una categoria analitica autonoma. Poiché non produce esclusione semantica diretta attraverso singole occorrenze lessicali, ma agisce in modo cumulativo e contestuale attraverso l'impiego di verbi e costruzioni associati all'agenzia e al potere, lo stesso non è stato incluso nei conteggi quantitativi dei termini discriminatori o non inclusivi[1]. Per questa ragione, la suddetta dimensione è stata analizzata esclusivamente mediante un approccio qualitativo. L'individuazione e la validazione del linguaggio mascolinizzante sono state effettuate tramite un'analisi manuale supportata dall'uso del *Gender Decoder* di Kat Matfield, impiegato come strumento di ausilio interpretativo e non come misura automatica[2].

Nell'ambito dell'analisi quantitativa, si è optato per la normalizzazione dei dati per 1.000 parole, adottata per ridurre la variabilità dovuta alla diversa lunghezza degli articoli, procedura comune negli studi di linguistica dei corpora che permette di attenuare l'impatto di articoli particolarmente brevi o lunghi (*cfr.* Biber *et. al.* 1998; McEnery & Hardie 2012; Gries, S. T 2009).

[1] Per ragioni di brevità e chiarezza espositiva, nei paragrafi successivi i termini "inclusivo", "biased", "discriminatorio" e "non inclusivo" verranno impiegati in modo intercambiabile.

[2] Questo strumento è stato sviluppato su ispirazione di un articolo di ricerca di Danielle Gaucher, Justin Friesen e Aaron C. Kay nel 2011, intitolato Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality, pubblicato sul Journal of Personality and Social Psychology (luglio 2011, 101, 1, pp. 109-128). Sebbene pensato come strumento di individuazione di un linguaggio discriminatorio negli annunci di lavoro, si è rivelato molto utile ai fini della presente analisi. *Cfr.* https://gender-decoder.katmatfield.com/

Sul piano operativo, è stato adottato un approccio misto, combinando strumenti quantitativi e qualitativi. Se da un lato, il conteggio delle occorrenze, unitamente alla *Sentiment Analysis* (condotta tramite il software Lingmotif)[3], ha permesso di rilevare la frequenza e la distribuzione di polarità positive, negative o neutre. Dall'altro, l'analisi semantica ha permesso di indagare in che modo il linguaggio concorra alla costruzione dei significati sul piano lessicale (termini discriminatori e non inclusivi), composizionale (combinazioni e contrasti semantici), pragmatico (presupposizioni e contesto) e sintattico (uso della voce attiva e passiva). Inoltre, l'applicazione della *semantic role labeling* (SRL) ha reso possibile osservare come agenti e pazienti venissero rappresentati, pur con i limiti derivanti dalla frequente soppressione dell'agente nei testi giornalistici.

Infine, tutti i risultati sono stati sottoposti a una validazione manuale del campione di articoli per verificare che le occorrenze individuate corrispondessero effettivamente a usi discriminatori o non inclusivi, ai fini di ridurre il rischio di falsi positivi e garantire maggiore solidità nelle interpretazioni.

3. Risultati

Dalla combinazione dell'approccio quantitativo e qualitativo al corpus emergono differenze particolarmente rilevanti tra le quattro testate selezionate, sia nella frequenza di termini connotati da bias di genere, sia nel tono complessivo dei testi.

In primo luogo, in linea con quanto osservato da Caliskan *et al.* (2022), i termini maschili individuati dagli autori del presente studio e da *Gender Decoder* tendono a comparire come verbi di agenzia, controllo e potere: "*lead*", "*dominate*", "*strong*". Nel corpus questi lemmi sono risultati particolarmente frequenti in riferimento all'atleta, rinforzando implicitamente l'associazione tra mascolinità, potere e successo (*cfr.* Figura 1).

[3] Lingmotif è una suite di analisi testuale multilingue avanzata che esamina i testi in ottica di *Sentiment Analysis*. Realizzata dal gruppo Tecnolengua dell'Università di Málaga, è in grado di analizzare sia contenuti generali sia domini specialistici. *Cfr.* https://ltl.uma.es/

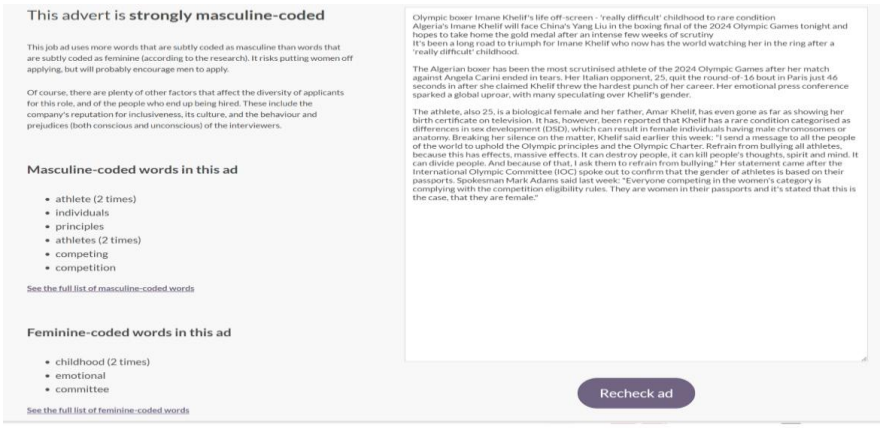

Figura 1- Esempio di analisi generata da Gender Decoder di Katmatfield[4].
Fonte: elaborazione propria

Dall'analisi degli articoli oggetto di studio, la frequenza più elevata di termini discriminatori sembra registrarsi nel *The New York Times*, mentre è il *Daily Mirror* a presentare il numero più basso di termini biased (Figura 2).

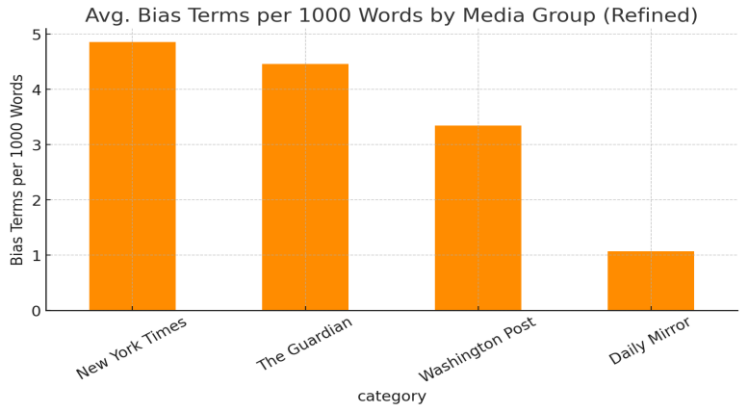

Figura 2- Termini discriminatori nelle 4 testate ogni 1000 parole.
Fonte: elaborazione propria

[4] *Cfr.* https://gender-decoder.katmatfield.com/

Come si evince anche dal confronto con normalizzazione per 1.000 parole, il ricorso a un tipo di linguaggio non appropriato è stato frequente anche nel caso del *The Guardian* (Tabella 1).

source	total_bias_terms	bias_terms_per_1000
TG- abandons ring	4	10.96
NYT- SHES A WOMAN	6	7.2
NYT- Im positive	11	6.6
WP- gender eligibility	8	4.99
WP olympic medal	5	3.8
DM- tears	5	1.64
TG- medal Im a woman	1	1.25
WP- Trump Imane	1	1.24
TG- semi-victory	1	1.17
DM- Olympic bosses	3	0.96
NYT- Furor	1	0.77
DM- Childhood	2	0.62

Tabella 2- Tot. termini biased per articolo vs tot. termini biased ogni 1000 parole[5].
Fonte: elaborazione propria

La *Sentiment Analysis* a livello frasale effettuata tramite la piattaforma Lingmotif ha, invece, evidenziato profili contrastanti, apparentemente divergenti dall'analisi di frequenza (Figura 3), in cui il *The Guardian* si distingue per il tono prevalentemente positivo e celebrativo il *The New York Times* per la distribuzione bilanciata tra positivo, negativo e neutro, il *Washington Post* per l'uso di un tono più neutro, mentre il *Daily Mirror* per la forte oscillazione tra polarità negative e sensazionalismo positivo.

[5] Per brevità, le quattro testate nella tabella sono indicate come di seguito: *The Guardian* (TG), *The New York Times* (NYT), *The Washington Post* (WP), *Daily Mirror* (DM).

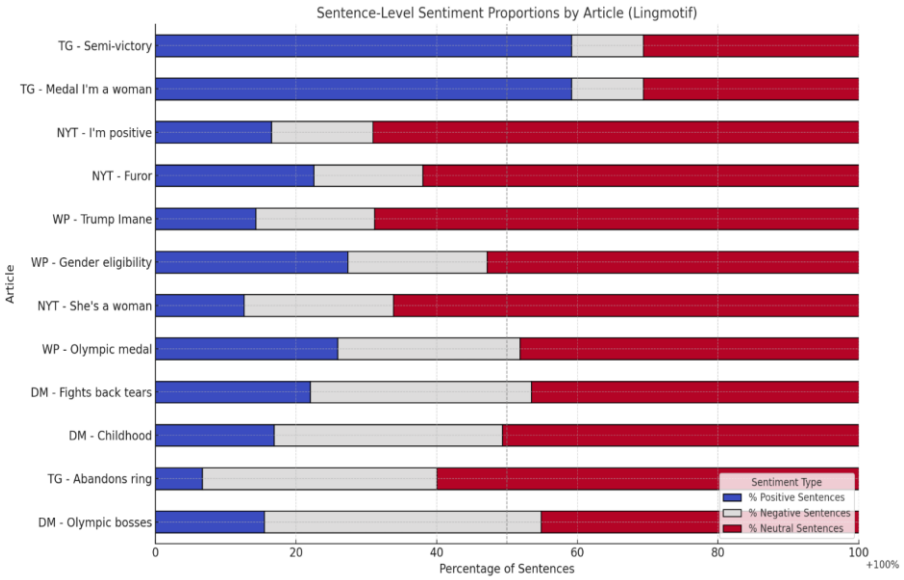

Figura 3- Sentiment Analysis a livello frasale realizzata con Lingmotif
Fonte: elaborazione propria

Tuttavia, tale dato risulta potenzialmente fuorviante se non incrociato con l'analisi semantica. Nel caso del *New York Times*, prevale infatti un linguaggio tecnico e scientifico che, pur presentandosi come neutro e oggettivo, tende a naturalizzare categorie biologiche e a ridurre la dimensione sociale e identitaria del genere, rientrando così, apparentemente, in una categoria di linguaggio non inclusivo. Come evidenziato in letteratura (Caliskan *et al.* 2022; Gaucher *et al.* 2011), infatti, questo tipo di registro discorsivo può produrre forme di non-inclusività implicita, poiché oscura l'agenzia dei soggetti coinvolti, limitando la possibilità di una lettura critica delle dinamiche sottese.

L'*Overall Sentiment Distribution* su sei categorie generato con *Lingmotif* (Figura 4) mostra, infatti, una distribuzione apparentemente equilibrata tra polarità positiva, negativa e neutra; tuttavia, l'analisi semantica rivela che tale equilibrio emotivo coesiste con scelte terminologiche e strutturali che contribuiscono a un'esclusione simbolica, rendendo il linguaggio formalmente neutro ma discorsivamente non inclusivo.

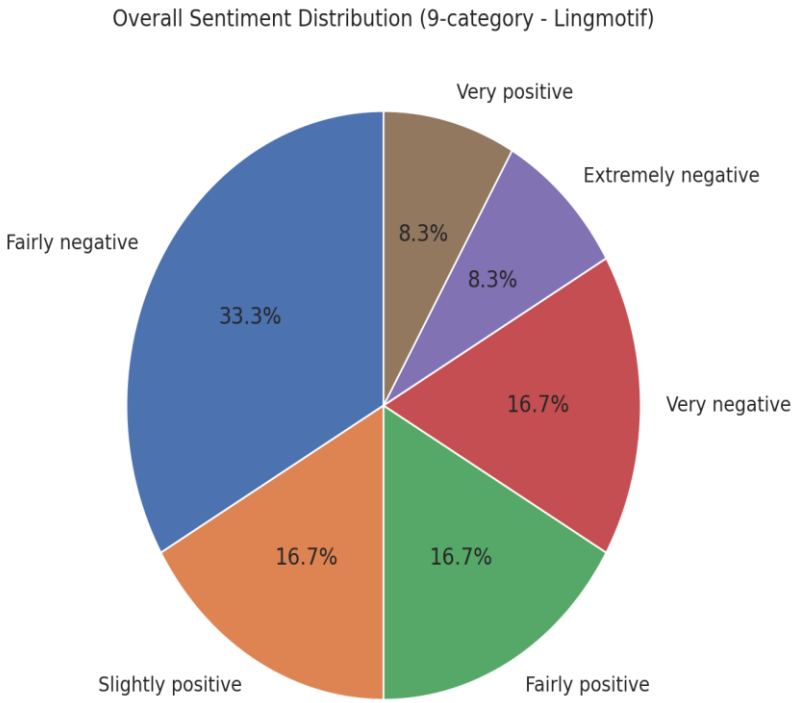

Figura 4- Overall Sentiment Distribution
Fonte: elaborazione propria

L'analisi del *sentiment* ha dunque fornito un quadro del tono del messaggio, ma non dei presupposti ideologici e metalinguistici del corpus preso in esame.

È proprio in questa direzione che si inserisce l'analisi semantico-qualitativa che ha infatti confermato l'esistenza di schemi discorsivi ricorrenti, quali:

- Costruzioni concessive (*"despite identifying as a woman"*) che insinuano dubbio
- Voce passiva (*"was disqualified"*) che oscura l'agente, contrapposta all'attiva (*"officials disqualified her"*)
- Lessico tecnicizzato (*"eligibility issue"*), che presenta conflitti come procedure burocratiche
- Metafore drammatizzanti (*"gender storm"*) che spettacolarizzano il caso

Nel complesso, i dati quantitativi e qualitativi mostrano che la natura del bias non dipende solo dalla frequenza, ma soprattutto dal modo in cui il linguaggio costruisce e distribuisce significati.

4. Discussione

I risultati ottenuti mostrano che se, da un lato, i dati quantitativi riflettono differenze nette nella frequenza di termini marcati (*es. real woman; male category; gender-row)* tra le testate; dall'altro, l'analisi semantica e quella del sentiment rivelano che tali differenze non possono essere lette in modo lineare, ma solo dalla combinazione di aspetti quali quantità, distribuzione, tono e struttura linguistica.

Un primo nodo interpretativo riguarda proprio il linguaggio discriminatorio che comprende espressioni che svalutano o delegittimano apertamente identità e soggetti (es. *real woman*). Il linguaggio non inclusivo, invece, si caratterizza per forme apparentemente neutre che, pur non offensive, escludono o invisibilizzano alcune identità (*es. the male category*). Pertanto, la corretta interpretazione della suddetta distinzione è essenziale per comprendere la natura del bias: la prima categoria produce un effetto diretto ed evidente, la seconda agisce in maniera più sottile e sistematica.

Da un punto di vista quantitativo, la normalizzazione per 1000 parole, ha permesso di comparare articoli di lunghezza molto diversa tra loro, la cui analisi avrebbe potuto condurre a risultati fuorvianti. Nello specifico, come mostrato dalla Figura 2, il *The New York Times* apparentemente registra la frequenza più alta di termini marcati, seguito dal *The Guardian*, mentre il *Daily Mirror* quella più bassa. Tuttavia, l'analisi semantica rivela che la natura del linguaggio è profondamente diversa: nel *NYT* prevale un lessico tecnico e scientifico, che rientra nella categoria del linguaggio non inclusivo e che normalizza una visione binaria dell'identità; nel *Mirror*, invece, si riscontrano meno occorrenze, ma formulate in modo apertamente discriminatorio, con espressioni che negano direttamente la legittimità dell'atleta o ne esaltano in modo smodato determinate caratteristiche.

Un ulteriore livello interpretativo riguarda il linguaggio mascolinizzante. In linea con quanto presentato in letteratura, termini come *lead*, *dominate* e *strong* ricorrono come verbi di agenzia, controllo e potere che, nel corpus oggetto di studio, sono stati impiegati con particolare frequenza per descrivere in riferimento all'atleta, rinforzando implicitamente l'associazione tra mascolinità, successo e autorità. Coerentemente con la letteratura precedente, pur non rientrando nella categoria del linguaggio discriminatorio esplicito, questo fenomeno rappresenta una forma di bias più sottile, riconducibile al linguaggio

non inclusivo, che contribuisce a rafforzare modelli di rappresentazione gerarchici.

La Sentiment Analysis applicata con l'ausilio di Lingmotif (Figura 3 e 4), ha aggiunto una dimensione ulteriore, ma con limiti significativi. La polarità positiva o negativa restituisce il tono complessivo di un testo, ma non i suoi presupposti ideologici. Di conseguenza, articoli che descrivono Khelif come "eroina" ottengono un punteggio positivo, pur contenendo espressioni non inclusive; al contrario, articoli critici verso le istituzioni risultano negativi, pur difendendo l'atleta (*was unfairly disqualified*). La Sentiment Analysis, quindi, fornisce indicazioni sulla percezione positiva o negativa del testo da parte di chi lo legge, ma non sull'effetto del testo stesso. Motivo per cui, se considerata isolatamente, l'analisi del sentiment rischia di restituire letture fuorvianti.

Infatti, l'analisi semantica ha permesso di evidenziare i processi discorsivi che veicolano i bias di genere. A titolo esemplificativo, l'uso della voce passiva (*was disqualified*), rispetto all'attiva (*officials disqualified her*), contribuisce a oscurare l'agente istituzionale, spostando l'attenzione dall'azione alla sua conseguenza. Le formule concessive (*despite identifying as a woman*) implicano un dubbio sull'autenticità dell'identità, mentre il ricorso a lessico tecnico-scientifico (*eligibility issue*) più ricorrente nel caso del New York Times e del *The Guardian* depoliticizza questioni sociali, trasformandole in cronaca burocratica. Viceversa, il ricorso a metafore drammatizzanti (gender storm), più comuni invece nel caso del *Daily Mirror* e del *The Washington Post*, tende a trasformare le stesse questioni in situazioni di conflitto sensazionalistico.

Dal punto di vista metodologico, questi risultati confermano l'importanza di un approccio combinato, in cui le frequenze numeriche, la l'analisi del sentiment e l'analisi semantica non possono essere considerate separatamente. Le prime forniscono dati comparativi, la seconda restituisce il tono percepito, la terza rivela i meccanismi linguistici più profondi. Solo integrando questi livelli si può comprendere perché il *NYT*, pur presentando la frequenza più alta di termini marcati, veicoli bias soprattutto attraverso linguaggio non inclusivo e mascolinizzante, mentre il *Mirror*, con meno occorrenze, ricorra a un linguaggio apertamente discriminatorio e spettacolarizzante.

Rilevante è stata anche la revisione manuale, che ha ridotto i falsi positivi, e l'uso di strumenti come la *Semantic Role Labeling (SRL)* che ha mostrato meccanismi di soppressione dell'agente, pur tenendo conto dei limiti derivanti dalla natura stessa del linguaggio giornalistico.

Conclusioni

I risultati dell'analisi mettono in luce la complessità del rapporto tra linguaggio, bias e rappresentazione mediatica, in cui ciascun livello metodologico mostra un aspetto diverso ma complementare.

In questo contesto, l'analisi linguistica svolta rappresenta un punto centrale del presente contributo. L'approccio adottato ha permesso di osservare come i bias di genere non emergano esclusivamente attraverso l'uso di termini esplicitamente discriminatori o non inclusivi, ma anche tramite scelte lessicali, sintattiche e discorsive all'apparenza neutre. In particolare, l'analisi semantica e pragmatica ha evidenziato il ruolo svolto dalla distribuzione dell'agenzia, dall'uso di registri tecnico-scientifici e dalla ricorrenza di specifici pattern verbali nella costruzione di rappresentazioni mediatrici del genere, mostrando come tali meccanismi possano contribuire a forme di esclusione simbolica difficilmente rilevabili attraverso il solo conteggio dei termini o l'analisi di polarità emotiva.

Per questo studio, una delle maggiori difficoltà è stata quella nell'identificare con precisione quali testate giornalistiche utilizzano effettivamente l'intelligenza artificiale nei loro processi produttivi. Infatti, i giornali identificati sono gli unici ad aver dichiarato l'uso dell'IA nei propri processi redazionali. Questa sfida metodologica rivela una lacuna critica nell'ecosistema informativo contemporaneo che è legato alla trasparenza nell'adozione di *tool* di IA da parte degli organi di informazione.

Lo studio esplorativo qui proposto è stato supportato anche da una revisione della letteratura scientifica, che ha dimostrato quanto l'impiego dell'intelligenza artificiale nel giornalismo si estenda ben oltre la semplice generazione automatica di testi, andando ad abbracciare un ventaglio di applicazioni che spaziano dalla revisione editoriale alla traduzione automatica, all'ottimizzazione SEO alla creazione di contenuti multimediali (Illescas Reinoso 2025: 2354-2375). Questa multidimensionalità dell'uso dell'IA complica significativamente l'identificazione e la valutazione del suo impatto sui contenuti informativi. Un articolo apparentemente scritto da un giornalista umano potrebbe essere stato sottoposto a processi di revisione automatica, tradotto in varie lingue ed ottimizzato per i motori di ricerca, attraverso strumenti AI, con l'aggiunta di altri contenuti informativi quali foto, video, audio, grafici, ecc. generati interamente o parzialmente tramite l'intelligenza artificiale (Verma 2024: 150-156).

La questione della trasparenza si complica ulteriormente quando si considera l'attuale panorama delle policy editoriali. Molte testate giornalistiche non forniscono dichiarazioni chiare e accessibili riguardo al loro utilizzo dell'IA, mentre quelle che lo fanno non sempre adattano le policy che possono essere poi verificate.

Un aspetto particolarmente critico che non bisogna sottovalutare è il ruolo dei collaboratori freelance nell'ecosistema giornalistico contemporaneo. La crescente dipendenza delle testate da contributi esterni introduce un ulteriore livello di complessità nella gestione dell'uso dell'IA. Tenendo conto che i freelance operano spesso in condizioni di precarietà economica e sotto pressioni temporali significative (Gutiérrez-Cuesta *et al.* 2022: 113-125), è possibile presumere che potrebbero ricorrere a strumenti di intelligenza artificiale per ottimizzare la loro produttività, indipendentemente dalle policy ufficiali delle testate per cui scrivono, generando così un paradosso in cui le testate giornalistiche possono avere policy chiare sull'uso dell'AI, ma l'implementazione pratica di tali norme diventa problematica quando una parte significativa dei contenuti viene prodotta da professionisti esterni che operano secondo logiche e vincoli diversi. La frammentazione del processo produttivo giornalistico rende quindi estremamente difficile garantire coerenza nell'applicazione di principi etici e metodologici relativi all'uso dell'intelligenza artificiale.

Inoltre, dall'analisi linguistica è emerso che un linguaggio formalmente oggettivo e scientifico, come quello riscontrato in alcune testate, non è necessariamente sinonimo di neutralità discorsiva. Al contrario, la combinazione tra un registro tecnico, una distribuzione bilanciata del sentiment e specifiche scelte semantiche può contribuire a naturalizzare categorie controverse, bias impliciti e, di conseguenza, a ridurre la complessità sociale dei fenomeni analizzati. Ciò conferma l'importanza di affiancare strumenti di analisi automatizzata a un esame linguistico e discorsivo accurato, in grado di cogliere forme di bias più sottili.

Per tali ragioni, i risultati di questo studio esplorativo hanno importanti implicazioni per la comprensione di come i bias di genere possano essere perpetuati e amplificati nell'attuale ecosistema informativo. Tuttavia, la difficoltà di identificare in modo affidabile l'uso dell'intelligenza artificiale nella produzione dei contenuti rende altrettanto complesso tracciare l'origine e i percorsi di diffusione dei bias e del linguaggio non inclusivo all'interno del discorso mediatico. In conclusione, emerge la necessità di sviluppare framework normativi e metodologici più sofisticati per affrontare le sfide poste dall'integrazione dell'IA nel giornalismo. È necessario che i media giornalistici siano vincolati tramite normative ad hoc per indurli ad essere più rigorosi e trasparenti sull'utilizzo dell'IA.

Per esempio, l'UE ha introdotto il *Digital Services Act* (European Commission 2024) e il *Digital Markets Act* (European Commission 2024) con lo scopo di creare un ambiente digitale più sicuro, equo e competitivo. Il DSA si concentra sui contenuti e sulle responsabilità degli intermediari online, imponendo loro di

rimuovere contenuti illegali, garantire la trasparenza della pubblicità e proteggere i diritti fondamentali degli utenti. Invece, il DMA si occupa della concorrenza nel mercato digitale, designando le grandi piattaforme online come *gatekeeper* e imponendo loro specifiche regole per prevenire pratiche commerciali sleali e favorire un mercato più aperto.

Ebbene, questi due regolamenti riguardano i giornali solo in maniera indiretta, nonostante - a livello normativo- ci sia stato un passo avanti importante con l'*AI Act* dell'Unione Europea (European Union 2024) che rappresenta la prima legislazione al mondo per regolamentare l'intelligenza artificiale in modo sistematico.

Tale regolamento classifica i sistemi *AI* in quattro categorie, secondo le quali, i sistemi a rischio inaccettabile vengono vietati completamente, quelli ad alto rischio devono rispettare standard rigorosi di trasparenza e supervisione umana, quelli a rischio limitato hanno obblighi più leggeri, e per finire, quelli a rischio minimo rimangono sostanzialmente liberi. Ciò significa che i sistemi *AI* generativi dovranno dichiarare quando il contenuto è artificiale per contrastare *deepfake* e disinformazione. Tuttavia, la normativa presenta limiti importanti dato che l'innovazione tecnologica evolve molto più rapidamente dei tempi legislativi e la stessa definizione di intelligenza artificiale rimane ambigua creando incertezze interpretative.

Lo studio ha rivelato che la questione dell'intelligenza artificiale nel giornalismo è molto complessa ed è estremamente difficile valutare con precisione l'impatto sulla qualità dell'informazione e sulla perpetuazione di bias discriminatori. Tuttavia, questa stessa complessità evidenzia l'importanza cruciale di continuare a investigare questi fenomeni dato che la democrazia dipende dalla qualità dell'informazione verso i cittadini.

La complessità del fenomeno analizzato, per eventuali studi futuri, richiede un approccio che tenga conto delle molteplici dimensioni coinvolte, non solo dei fattori tecnologici e linguistici, ma anche di quelli sociali e politici. La letteratura esistente ha esplorato separatamente molti di questi aspetti, ma rimane una lacuna significativa nell'analisi integrata di come l'intelligenza artificiale, le *fake news* e i *bias* di genere interagiscono nella costruzione di narrazioni politicamente strumentalizzate. Questa lacuna è particolarmente evidente nell'analisi di casi specifici che illustrano concretamente questi meccanismi, come ad esempio quello di questa ricerca sul caso di Imane Khelif che potrebbe essere maggiormente approfondito ampliando il campione delle testate giornalistiche da analizzare, ed estendendo lo studio non limitandosi alla sola stampa anglosassone e statunitense.

Bibliografia

AMERICAN PSYCHOLOGICAL ASSOCIATION, *Publication manual of the American Psychological Association*. Washington, DC: APA, 2020 (7ª ed.).

AL-ASADI, M. A.; TASDEMIR, E., "Artificial intelligence and gender bias in social media content moderation". In: *Journal of Digital Ethics*, 15, 3, 2022, pp. 45-62. Accessed at: https://doi.org/10.1007/s11948-022-00389-7

ALLCOTT, H.; GENTZKOW, M., "Social media and fake news in the 2016 election". In: *Journal of Economic Perspectives*, 31, 2, 2017, pp. 211-236. Accessed at: https://doi.org/10.1257/jep.31.2.211

BAKER, P., *Using corpora in discourse analysis*. London: Continuum, 2006.

BENKLER, Y.; FARIS, R.; ROBERTS, H., *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford: Oxford University Press, 2018. Accessed at: https://doi.org/10.1093/oso/9780190923624.001.0001

BIJU, A.; GAYATHRI, E. S., "Gender bias amplification in AI-driven content moderation systems". In: *AI & Society*, 38, 4, 2023, pp. 1567-1582. Accessed at: https://doi.org/10.1007/s00146-022-01478-3

BOLUKBASI, T.; CHANG, K. W.; ZOU, J. Y.; SALIGRAMA, V.; KALAI, A. T., "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings". In: *Advances in Neural Information Processing Systems*, 29, 2016, pp. 4349-4357. Accessed at: https://doi.org/10.5555/3157382.3157584

BRIK, S., *The Construction of Narratives Around Postcolonial Algerian Identity: Examining Imane Khelif Journey During Paris Olympic Games 2024*. Tesi di dottorato, University of Martyr Sheikh Larbi Tebessi Tebessa, 2025. Accessed at: http//localhost:8080/jspui/handle/123456789/12814

CALISKAN, A.; BRYSON, J. J.; NARAYANAN, A., "Semantics derived automatically from language corpora contain human-like biases". In: *Science*, 356, 6334, 2017, pp. 183-186. Accessed at: https://doi.org/10.1126/science.aal4230

CHESNEY, B.; CITRON, D., "Deep fakes: A looming challenge for privacy, democracy, and national security". In: *California Law Review*, 107, 6, 2019, pp. 1753-1820. Accessed at: https://doi.org/10.15779/Z38RV0D15J

EHRLICH, S., *Representing rape: Language and sexual consent*. London: Routledge, 2001.

ESSER, F.; NEUBERGER, C., "Realizing the democratic functions of journalism in the digital age: New alliances and a return to old values". In: *Journalism*, 20, 1, 2018, pp. 194-197. Accessed at: https://doi.org/10.1177/1464884918807067

EUROPEAN COMMISSION, *The Digital Markets Act*, 2024. Accessed at: https://digital-markets-act.ec.europa.eu/index_en

——, *The Digital Services Act*, 2024. Accessed at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

EUROPEAN UNION, *Regulation 2024/1689 of the European Parliament and of the Council*, 2024. Accessed at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

EUROPEAN INSTITUTE FOR GENDER EQUALITY, *Gender-sensitive communication*. Luxembourg: Publications Office of the European Union, 2016.

FAIRCLOUGH, N., *Critical discourse analysis: The critical study of language*. London: Longman, 1995.

FLORIDI, L.; COWLS, J.; BELTRAMETTI, M.; CHATILA, R.; CHAZERAND, P.; DIGNUM, V.; VAYENA, E., "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations". In: *Minds and Machines*, 30, 4, 2020, pp. 689-707. Accessed at: https://doi.org/10.1007/s11023-018-9482-5

FREELON, D.; BOSSETTA, M.; WELLS, C.; LUKITO, J.; XIA, Y.; ADAMS, K., "Black trolls' matter: Racial and ideological asymmetries in social media disinformation". In: *Social Science Computer Review*, 38, 5, 2020, pp. 621-637. Accessed at: https://doi.org/10.1177/0894439319864157

GAUCHER, D.; FRIESEN, J.; KAY, A. C., "Evidence that gendered wording in job advertisements exists and sustains gender inequality". In: *Journal of Personality and Social Psychology*, 101, 1, 2011, pp. 109-128, Accessed at: https://doi.org/10.1037/a0022530

GUESS, A.; NAGLER, J.; TUCKER, J., "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". In: *Science Advances*, 5, 1, 2019. Accessed at: https://doi.org/10.1126/sciadv.aau4586

GUTIÉRREZ-CUESTA, J. J.; VINK LARRUSKAIN, N.; CANTALAPIEDRA GONZÁLEZ, M. J., "La precariedad, obstáculo para la calidad periodística: estudio de caso". In: *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, 35, 2022, pp. 113-125. Accessed at: https://doi.org/10.31921/doxacom.n35a1588

HOWARD, P. N.; BRADSHAW, S., "Junk news and bots during the U.S. election: What were Michigan voters sharing over Twitter?". In: *Computational Propaganda Research Project Working Paper*, 2017. Accessed at: https://doi.org/10.2139/ssrn.3088901

ILLESCAS REINOSO, D.; PALACIOS, A. G.; ORTIZ VIZUETE, F., "La inteligencia artificial en el periodismo: herramientas y aplicaciones". In: *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 6, 1, 2025, pp. 2354-2375. Accessed at: https://doi.org/10.56712/latam.v6i1.3503

LAZER, D. M.; BAUM, M. A.; BENKLER, Y.; BERINSKY, A. J.; GREENHILL, K. M.; MENCZER, F.; ZITTRAIN, J. L., "The science of fake news". In: *Science*, 359, 6380, 2018, pp. 1094-1096, Accessed at: https://doi.org/10.1126/science.aao2998

LOPEZ, P., "Bias does not equal bias: A socio-technical typology of bias in data-based algorithmic systems". In: *Policy & Interne*t, 2021. Department of Legal Philosophy, University of Vienna. Accessed at: https://doi.org/10.14763/2021.4.1598

MARCHETTI, R., "Gender representation in AI-generated media content: A critical analysis". In: *Digital Media Studies Quarterly*, 22, 4, 2019, pp. 78-95. Accessed at: https://doi.org/10.1080/13691183.2019.1623456

MARWICK, A.; LEWIS, R., *Media manipulation and disinformation online*. Data & Society Research Institute, 2017. Accessed at: https://doi.org/10.2139/ssrn.2958111

MINUCCI, S.; THOMPSON, K.; RODRIGUEZ, L., "Artificial intelligence and the perpetuation of gender stereotypes in digital media". In: *Computers in Human Behavior*, 128, 2022, pp. 107-115. Accessed at: https://doi.org/10.1016/j.chb.2021.107115

NAKOV, P.; DA SAN MARTINO, G., "Fake News, Disinformation, Propaganda, and Media Bias". In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21)*. New York: Association for Computing Machinery, 2021, pp. 4862-4865. Accessed at: https://doi.org/10.1145/3459637.3482026

NAZAR, R.; BUSTAM, M. R., "Gender bias in artificial intelligence systems: Implications for digital communication". In: *International Journal of Human-Computer Studies*, 143, 2020, pp. 102-118. Accessed at: https://doi.org/10.1016/j.ijhcs.2020.102567

NOBLE, S. U., *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press, 2018. Accessed at: https://doi.org/10.2307/j.ctt1pwt9w5

PARCU, P. L., "New digital threats to media pluralism in the information age". In: *Competition and Regulation in Network Industries*, 21, 2, 2019, pp. 91-109. Accessed at: https://doi.org/10.1177/1783591719886101

PÉREZ-ROSAS, V.; KLEINBERG, B.; LEFEVRE, A.; MIHALCEA, R., "Automatic detection of fake news". In: *Proceedings of the 27th International Conference on*

Computational Linguistics, 2018, pp. 3391-3401. Accessed at: https://doi.org/10.18653/v1/C18-1287

ROTH, Y.; PICKLES, N., "Updating our approach to misleading information". In: *Twitter Blog*, 2020. Accessed at: https://doi.org/10.48550/arXiv.2011.14751

SHOR, E.; VAN DE RIJT, A.; FOTOUHI, B., "A large-scale test of gender bias in the media". In: *Sociological Science*, 6, 2019, pp. 526-550. Accessed at: https://doi.org/10.15195/v6.a20

SHU, K.; SLIVA, A.; WANG, S.; TANG, J.; LIU, H., "Fake news detection on social media: A data mining perspective". In: *ACM SIGKDD Explorations Newsletter*, 19, 1, 2017, pp. 22-36. Accessed at: https://doi.org/10.1145/3137597.3137600

STAHLBERG, D.; BRAUN, F.; IRMEN, L.; SCZESNY, S., "Representation of the sexes in language". In: *Social Communication*, Psychology Press, 2007, pp. 163-187.

STUBBS, M., *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell, 2001.

SUNSTEIN, C. R., *#Republic: Divided democracy in the age of social media*. Princeton: Princeton University Press, 2017. Accessed at: https://doi.org/10.1515/9781400884711

TANDOC JR, E. C.; LIM, Z. W.; LING, R., "Defining 'fake news'. A typology of scholarly definitions". In: *Digital Journalism*, 6, 2, 2018, pp. 137-153. Accessed at: https://doi.org/10.1080/21670811.2017.1360143

TUCKER, J. A.; GUESS, A.; BARBERA, P.; VACCARI, C.; SIEGEL, A.; SANOVICH, S.; NYHAN, B., "Social media, political polarization, and political disinformation: A review of the scientific literature". In: *Political polarization, and political disinformation*, 2018, pp. 1-95. Accessed at: https://doi.org/10.2139/ssrn.3144139

UNESCO, *Guidelines on gender-neutral language in English*. Paris: United Nations Educational, Scientific and Cultural Organization, 2019.

VERMA, D., "Impact of Artificial Intelligence on Journalism: A Comprehensive Review of AI in Journalism". In: *Journal of Communication and Management*, 3, 2, 2024, pp. 150-156. Accessed at: https://doi.org/10.58966/JCM20243212

VOSOUGHI, S.; ROY, D.; ARAL, S., "The spread of true and false news online". In: *Science*, 359, 6380, 2018, pp. 1146-1151. Accessed at: https://doi.org/10.1126/science.aap9559

WEBER-WULFF, D.; ANOHINA-NAUMECA, A.; BJELOBABA, S.; FOLTÝNEK, T.; GUERRERO-DIB, J.; POPOOLA, O.; SIGUT, P.; WADDINGTON, L., "Testing of detection tools for AI-generated text". In: *International Journal for Educational*

Integrity, 19, 2023, pp. 1-39. Accessed at: https://doi.org/10.1007/s40979-023-00146-z

WEIDINGER, L.; UESATO, J.; RAO, R.; GRANT, E.; STANLEY, K.; BALDWIN, R.; COX, S.; WHITE, A.; MILANO, S.; STEINHARDT, J.; CHRISTIANO, P., "Ethical and social risks of harm from language models". In: *arXiv preprint*, 2021. Accessed at: https://arxiv.org/abs/2112.04359