

## POLITICAMENTE CORRETTO E SOCIAL MEDIA: UNO SPETTRO CENSORIO?

DOI: 10.7413/18281567266

di **Maria Rosaria Vitale**

Università degli Studi di Urbino Carlo Bo

### **Politically correctness and social media: a censorious spectre?**

#### *Abstract*

This essay examines the concerns surrounding political correctness and its perceived impact on freedom of speech, particularly in social media contexts. It analyzes how algorithmic content moderation shapes this debate. Given the claim that ‘we cannot say anything anymore’, I investigate the possible implications of the persistent criticism directed against political correctness.

**Keywords:** political correctness, social media, content moderation, freedom of speech, censorship.

‘Al giorno d’oggi non si può più dire nulla’. Questa è un’affermazione che è divenuta alquanto frequente *anche* nei social media, quali spazi digitali che pongono condizioni di espressione e interconnessione per la propria utenza. Sembra, quindi, che un’impossibilità di esprimersi sia rivendicata, esprimendosi. A mio avviso, l’affermazione presenta due dimensioni: una temporale e una spaziale. Prima ci si poteva esprimere, adesso non più. Il diritto di libertà di espressione subisce dei contraccolpi in società liberal-democratiche. La prima dimensione è, forse, più facilmente discernibile della seconda. Tuttavia, ritengo che la localizzazione di un’impossibilità di esprimersi in società liberal-democratiche, rispetto ad altre, possa essere assunta ed in quanto il discorso pubblico di queste include l’affermazione in questione. In un’intersezione delle due dimensioni, il cosiddetto politicamente corretto (PC) viene considerato ciò da cui muove tutto.

Se il PC fa sì che non si possa più dire nulla, esso non vanta una definizione unanimemente riconosciuta. Tale mancanza si ritrova nella letteratura scientifica e non solo. Cosa certa è che il dibattito sul PC resiste dalla fine degli anni Ottanta, così come ha un suo luogo originario, gli Stati Uniti, riuscendo a migrare altrove. Ci si può chiedere perché e con un'attenzione ai social media che sono stati creati (negli Stati Uniti) e nonostante il PC. O forse tali spazi digitali sono stati contaminati dal PC, sin dalla loro creazione – *ergo* spazi di espressione e interconnessione non particolarmente permissivisti.

Alla luce dell'acceso dibattito sul PC (scientifico, politico, nei e tramite i media tradizionali e i social media), propongo una rimodulazione dell'affermazione con cui si è aperto questo lavoro: 'Al giorno d'oggi non si può più dire nulla?'. Occorre, comunque, riflettere su cosa si intende variamente per PC, in modo tale problematizzare l'inclusività che ritengo insita nella domanda appena presentata.

### Disvelamenti

Affinché ci si muova entro il dibattito, Cameron presenta uno scenario fittizio, in cui antropologi marziani indagano il PC quale fenomeno delle società occidentali<sup>1</sup>. In una preliminare considerazione della letteratura umana anglofona, tale dibattito risulta essere caratterizzato per lo più da posizioni strenuamente critiche. Sono quest'ultime a ricondurre il PC ad un potente movimento politico che dalle università si è espanso nelle società di riferimento<sup>2</sup>. Chi presumibilmente vi aderisce (femministe, anti-razzisti e multiculturalisti) nega sia di farne parte, sia che un tale movimento di per sé esista<sup>3</sup>. Più nello specifico, tre sono i principali obiettivi imputati al PC:

---

\* Lì dove non menzionato, le traduzioni sono dell'Autrice.

<sup>1</sup> D. Cameron, *Verbal Hygiene*, Routledge, London-New York 1995.

<sup>2</sup> Cfr. D. D'Souza, *Illiberal Education: The Politics of Race and Sex on Campus*, Vintage, New York 1992; M. Friedman – J. Narvison, *Political Correctness. For and Against*, Rowman & Littlefield Publishers, Boston-Lanham-London 1995; K. Allan – K. Burrige, *The Language of Political Correctness*, in Id. *Forbidden Words. Taboo and the Censoring of Language*, Cambridge University Press, Cambridge 2006; G. Hughes, *Political Correctness. A History of Semantics and Culture*, Wiley-Blackwell, Oxford 2010.

<sup>3</sup> "Coloro che non fanno mistero del loro sostegno alle *affirmative action*, al linguaggio non-discriminatorio e ai curricula multiculturali, non fanno riferimento collettivamente a tali posizioni come 'correttezza politica': sono solo i loro oppositori a usare questa espressione in questo modo". Cameron, *Verbal Hygiene*, cit., p. 125.

trattamenti preferenziali ai membri di certi gruppi sociali (per esempio, donne, minoranze etniche) nelle scuole e nelle università; la costruzione di programmi educativi in cui le idee tradizionali di patrimonio culturale e di eccellenza artistica sono sostituite da un' enfasi sui contributi culturali non occidentali, non bianchi e femminili; e la prescrizione del tipo di linguaggio che può o non può essere utilizzato per parlare delle differenze tra gli esseri umani, in particolare delle differenze di genere e razziali/etiche<sup>4</sup>.

Un *fil rouge* rende conto dei tre obiettivi, e che si tratti di *affirmative action*, revisione dei curricula, controllo del linguaggio. Faccio qui riferimento all'identità ascrivibile altrui, analizzabile come tale tramite le caratteristiche ascrivibili dei soggetti – origine razziale, etnica, nazionale, religione, orientamento sessuale, genere, disabilità. In una selezione, intendo focalizzarmi sul terzo obiettivo, con la domanda sull'impossibilità di esprimersi che necessita di un'ulteriore rimodulazione: 'Al giorno d'oggi non si può più dire nulla su specifici soggetti e specifiche questioni, senza conseguenze nefaste o quasi?'. Per quanto chiunque sia tale anche alla luce delle proprie caratteristiche ascrivibili, la questione del PC permette di indagare ciò che viene detto a soggetti membri di gruppi minoritari, e di questioni che li riguardano direttamente e trasversalmente<sup>5</sup>. Le conseguenze prima citate si strutturano, specificamente, nella forma di accuse di -ismi e -fobie, riconducibili alle caratteristiche ascrivibili. Chi utilizza il maschile generico, è sessista. Chi ridicolizza coloro che definiscono il *coffee without milk*, invece di *black coffee*, razzializza il caffè. Chi è scettico sulla scelta di un interprete non-bianco per il ruolo di Macbeth, è razzista.

Alla luce del *fil rouge* ascrivibile, Aly e Simpson propongono una definizione di PC, trattandosi di "un insieme di pratiche sociali che comportano l'evitamento o il controllo di comportamenti – solitamente espressioni – che sono considerati denigranti nei confronti di soggetti appartenenti a gruppi sociali

---

<sup>4</sup> *Ivi*, pp. 124-125.

<sup>5</sup> Si veda R. T. Lakoff "Political Correctness" and Hate Speech, in Id. *The Language War*, University of California Press, Berkeley-Los Angeles-London, 2000; M. Talbot, *Political Correctness and Freedom of Speech*, in M. Hellinger – A. Pauwels (a cura di), *Handbook of Language and Communication: Diversity and Change*, De Gruyter, Berlin 2007; R. Sparrow, *Talking Sense about Political Correctness*, in «Journal of Australian Studies», 73/1, 2002.

subordinati”<sup>6</sup>. Presento qui una puntualizzazione. Le espressioni che potrebbero essere evitate o controllate non sono da sé illecite. Ciò vale certamente per il sistema normativo statunitense: il diritto di libertà di espressione vanta una protezione costituzionale ampia (e più ampia in relazione ad altri sistemi normativi statali), includendovi il c.d. *hate speech*, quale categoria di espressioni dirette a soggetti individuati per le proprie caratteristiche ascrivibili<sup>7</sup>. Questo è un aspetto curioso, nella misura in cui il PC – originariamente statunitense – sia emerso lì dove una libertà di espressione è legittimamente esercitabile al di là della prospettiva di senso di chi si esprime. Che quest’ultima sia ascrivibilmente - ista o - fobica, non è rilevante. “Il divieto costituzionale americano di *viewpoint discrimination* impedisce di trattare chi chiama gli altri ‘negri’ in modo diverso da chi chiama gli altri ‘figli di puttana’, e impedisce di trattare chi porta bandiere naziste in modo diverso da chi brucia quelle americane”<sup>8</sup>. Di contro e nella questione del PC, la prospettiva di senso (la c.d. *viewpoint*) - ista e - fobica diventa esattamente centrale. In sistemi normativi statali non-statunitensi, è riscontrabile una (opposta) regolamentazione della libertà di espressione nella forma dello *hate speech* alla luce della prospettiva di senso dei soggetti<sup>9</sup>. Tuttavia, l’esprimersi sulle caratteristiche ascrivibili altrui non si pone da sé come illecito. Ciò vale nell’utilizzo del maschile generico, di un termine usuale per il caffè e nell’insoddisfazione circa l’adattamento cinematografico di una tragedia shakespeariana. A questo punto, se le accuse di - ismi e - fobie sono mosse in un contesto statale piuttosto che in un altro (il PC è riuscito a migrare), esse implicano un’attenzione zelante che ogni soggetto dovrebbe porre su quanto intende esprimere. Con Cameron, i critici del PC “sono sinceramente disorientati dal fatto che donne o membri di gruppi etnici minoritari si ostinino a leggere connotazioni così bizzarre in parole perfettamente innocenti, il cui significato dovrebbe essere sicuramente chiaro a tutti”<sup>10</sup>. Ci si può, quindi, chiedere se al disorientamento occorra rispondere continuando a esprimersi come si è

---

<sup>6</sup> W. Aly – R. M. Simpson, *Political Correctness Gone Viral*, in C. Fox and J. Saunders (a cura di), *Free Speech and Media Ethics*, Routledge, London 2019, p. 125.

<sup>7</sup> Si veda J. Weinstein, *Hate Speech, Pornography, and the Radical Attack on Free Speech Doctrine*, Westview Press, Boulder 1999.

<sup>8</sup> F. Schauer, *The American First Amendment*, in M. Ignatieff (a cura di), *American Exceptionalism and Human Rights*, Princeton University Press, Princeton-London, p. 38.

<sup>9</sup> Si veda E. Heinze, *Viewpoint Absolutism and Hate Speech*, in « The Modern Law Review », 69/4, 2006; A. Brown – A. Sinclair, *The Politics of Hate Speech Laws*, Routledge, London-New York 2020.

<sup>10</sup> Cameron, *Verbal Hygiene*, cit. p. 163.

sempre fatto, o adattandosi a eufemismi, neologismi, sostituzioni, revisioni. In relazione a quest'ultimo scenario, il PC fa qualcosa, agendo sia preventivamente a ciò che si esprime (controbilanciando il rischio di accuse varie), sia successivamente (strutturando le condizioni per accuse ampie). Seguono degli esempi attribuiti al PC: l'asterisco nella lingua italiana (informale) per indicare il genere neutro, una donna *chairperson* o *chair*, sindaca o architetta, un soggetto disabile definito *phisically-challenged*, Edoardo VI Tudor che è nero e omosessuale in una serie tv, la formula *studentesse e studenti* in documenti amministrativi di alcuni atenei italiani. Si tratta di qualcosa di ridicolo o ridondante? Altri esempi: il *deplatforming* nei social media generalisti di alcuni sostenitori di Donald J. Trump, il *noplatforming* di relatori disinvitati ad eventi e convegni programmati in università, a seguito di rimostranze di una parte del corpo studentesco e docente, la ristampa di alcuni libri di Robert Dahl con modifiche rispetto alle versioni originali. Occorre qui sorridere o esserne preoccupati?

L'attenzione zelante, prima presentata, si collega ad un altrui e pervasivo leggere tra le righe che sembra porsi come un'opera di disvelamento. Il *corretto* nel PC potrebbe essere interpretato in termini di tendenza verso un adeguato esprimersi. Ipotizzando uno scenario, *'L'utilizzo del maschile generico è un residuo sessista resistente. Occorre essere ben consapevoli che una diversificazione di genere di ciò di cui si parla è plausibile'*; *'Così come mi esprimo non è sessista, ed io non sono sessista. Sei solo tu ad arrivare a questa conclusione'*. Basta adattare quanto presentato ad altre questioni che, esplicitamente o implicitamente, prendono in causa le caratteristiche ascrittive. All'esprimersi in una certa modalità corrisponde ciò che alternativamente viene rivendicato come adeguato. Da qui, l'elemento che può essere di scherno o di preoccupazione muove dall'attribuzione del PC (movimento politico o insieme di pratiche sociali) come proprio di alcuni soggetti e che si pone a danno di altri. I primi vengono definiti, in una buona parte del dibattito, quali radicali. I secondi sono generalmente coloro che divengono oggetto di una limitazione della libertà di espressione, consapevoli o meno. Avremo così, da una parte, i detentori dell'individuazione delle condizioni per il riconoscimento di -ismi e -fobie, e dall'altra, chi ne viene accusato. Al riguardo, la mia ipotesi circa l'inclusività insita nella domanda *'Al giorno d'oggi non si può più dire nulla?'* si caratterizza per un parallelo disvelamento attuato dai critici del PC. Il *nulla* diventa comprensivo di qualunque cosa, con i radicali che vengono considerati coloro che mettono in discussione ciò che lecito dire, leggere, guardare, in una smodata ricerca di odio e discriminazione. Di tale domanda ho proposto una

rimodulazione, specificando l'impossibilità di esprimersi su certi soggetti e certe questioni. Anche in questo caso, il rivendicato *nulla* non chiarifica, rimanendo assai comprensivo. Se il PC consiste in un "gergo del divisionismo nevrotico"<sup>11</sup> e un mero "galateo politico, non politica vera"<sup>12</sup>, Hughes, insieme ai critici come lui, deride lo stesso PC. Difatti, "nessuna sostituzione di parole è in grado di ridurre il tasso di intolleranza presente in questa o qualunque altra società"<sup>13</sup>. Il disvelamento riguarda qui la considerazione secondo cui ogni sforzo in vista della *correttezza* sia futile, sebbene destabilizzante. Una divisione si realizza, con un linguaggio inclusivo che è escludente, se quanto si esprime è lecito. Sono quindi i confini di un diritto di libertà di espressione ad essere divenuti assai labili. Ci si può esprimere, ma non tanto alle sole condizioni poste dal proprio sistema normativo. Il PC definisce le sue (parallele) condizioni, con i soggetti esposti ad un altrui continuo e pervasivo leggere tra le righe. Con Origgi e in un adattamento<sup>14</sup>, potrei chiedermi 'come mi vedo vista?', dato l'utilizzo del maschile generico in queste pagine. Ritengo, comunque, che tale domanda innervi la questione del PC, in quanto l'attenzione che un soggetto pone su ciò che intende esprimere muove dall'attenzione altrui che gli viene diretta.

Il mero riferimento alle accuse di - ismi e - fobie non rende conto esaustivamente del dibattito sul PC, e nella misura in cui i critici hanno proposto svariate definizioni: che si tratti di un nuovo maccartismo, fondamentalismo, illiberalismo, autoritarismo, totalitarismo, di una realizzata *thought police* orwelliana<sup>15</sup>, l'imputazione di una censura è centrale<sup>16</sup>; la legittimità di un diritto di libertà di espressione viene assunto come un indicatore fondamentale della tenuta di società liberal-democratiche. Se lo stesso dibattito è emerso in tempi in cui i social media non erano stati creati, l'avvento di spazi digitali di espressione e interconnessione può essere oggetto di una problematizzazione. Affermare che i social media non siano mai stati propriamente permissivisti è

---

<sup>11</sup> R. Hughes, *La cultura del piagnisteo. La saga del politicamente corretto*, tr. it. di M. Antonielli, Adelphi, Milano 1994, p. 40. Cfr. testo in lingua originale, *Culture of Complaint. The Fraying of America*, Oxford University Press, New York-London 1993.

<sup>12</sup> Hughes, *La cultura del piagnisteo*, cit., p. 41.

<sup>13</sup> *Ibidem*.

<sup>14</sup> G. Origgi, *La reputazione. Chi dice cosa di chi*, EGEA-Università Bocconi Editore, Milano 2016.

<sup>15</sup> Si veda nota n. 2.

<sup>16</sup> Cfr. R. Post, *Censorship and Silencing: Practices of Cultural Regulation* (a cura di), Getty Publications, Los Angeles 1998; I. Peleg, *Patterns of Censorship Around the World* (a cura di), Routledge, New York-Abingdon 2019.

inesatto. Difatti, le critiche sono state frequenti circa un'indifferenza nei confronti di contenuti *user-generated* illeciti, al confine tra illecito e lecito, o comunque problematici. Ciò non implica, tuttavia, che i social media siano degli spazi che non (auto-)regolamentano i contenuti della propria utenza: la *content moderation* è “essenziale, costitutiva e definitoria”<sup>17</sup>, rendendo i social media esattamente quel che sono; non tutti i contenuti sono pubblicabili e nessun contenuto è visibile all'utenza in modo randomico. La moderazione pone così condizioni di espressione e interconnessione, le quali sono state ridefinite negli anni, fino ad arrivare ad una rivendicazione secondo cui i social media siano diventati censori. E la causa di ciò viene attribuita al PC.

A questo punto, se è plausibile che l'utenza si esprima anche sulle caratteristiche ascrivite altrui (esplicitamente e implicitamente), intendo strutturare l'analisi che segue tramite due questioni. La prima concerne l'algorithmica individuazione e valutazione dei contenuti tramite *content moderation*, tale per cui i social media, tra le cose, li eliminano o ne riducono la visibilità. Nella seconda, l'interconnessione viene assunta in quanto fa sì che ogni utente possa essere attento su chiunque altro – su come si esprime e come è adeguato esprimersi.

### **Interferenze**

La *content moderation* è multiforme<sup>18</sup>. Essa consiste nella cura dei contenuti *user-generated* (testi, video, audio, immagini). Questi potranno essere visibili, amplificati e de-amplificati ad un soggetto-utente piuttosto che ad un altro, così come eliminati perché non conformi alle regole che i social media definiscono per la propria utenza. La moderazione ha uno scopo ben preciso e che rende conto dei social media quali *corporation*: i flussi di dati generano profitti e l'esperienza digitale dell'utenza deve essere resa il più soddisfacente possibile. Alla luce della quantità considerevole dei contenuti dell'utenza, la moderazione sarà preponderatamente algorithmica rispetto a quella affidata ai team di

---

<sup>17</sup> T. Gillespie, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, Yale University Press, New Haven-London 2018, p. 21.

<sup>18</sup> Si veda T. Gillespie, *The Custodians of the Internet*, op. cit.; T. Gillespie, *Content Moderation, AI, and the Question of Scale*, in «Big Data & Society», 7/2, 2020; E. Douek, *Content Moderation as Systems Thinking*, in «Harvard Law Review», 136/2, 2022; S. T. Roberts, *Behind the Screen. Content Moderations in the Shadows of Social Media*, Yale University Press, New Haven-London 2019; L. Edelson, *Content Moderation in Practice*, in K. Langvardt – J. Hurwitz (a cura di), *Media and Society After Technological Disruption*, Cambridge University Press, Cambridge 2024.

moderatori ‘umani’. La valutazione di quest’ultimi, comunque, è tale grazie ad una precedente individuazione algoritmica e quindi automatizzata dei contenuti.

In una selezione, intendo concentrarmi sul prima accennato *deplatforming*. Il riferimento è stato all’eliminazione degli account di alcuni sostenitori di Trump, durante il suo primo mandato presidenziale e successivamente ai disordini di Capitol Hill. In relazione a quest’ultimi, gli stessi account di Trump sono stati sospesi da più piattaforme. Si tratta di casi ben specifici, ma indicativi di un’impossibilità di espressione che grava su alcuni soggetti. Il PC viene posto lungo l’asse politico, divenendo proprio di una non ben definita sinistra, a svantaggio di una non ben definita destra. O meglio e tra le varie interpretazioni, la prima sarà radicale, conformista, la seconda controversa e vittima di una *correttezza* imperante, e ciò dipende dalla prospettiva da cui si guarda il PC. Una censura dei social media generalisti è rivendicata, alla luce di una limitazione della libertà di espressione che viene definita faziosa, appunto, lungo l’asse politico. Lo stesso posizionamento era stato, comunque, individuato ancor prima dell’avvento dei social media.

Gorissen considera il *deplatforming* nei termini di un processo multi-livello di moderazione attiva che costituisce una “fondamentale minaccia all’esistenza di movimenti e *community*”<sup>19</sup>, non trattandosi della mera eliminazione di alcuni account. Un primo livello micro è quello dell’*individual deplatforming* e il caso di Trump è intuitivo. La sospensione dell’account Twitter di Trump, allora presidente uscente, è stata giustificata per la mancata conformità alla *policy* concernente la glorificazione della violenza. La decisione del *deplatforming* di Trump è stata giustificata tramite un’argomentazione resa pubblica nella sezione *Blog* dell’allora Twitter, cosa che non è riscontrabile per il *deplatforming* di altri utenti – il profilo di Trump non era semplicemente uno tra i tanti; difatti e per chiunque altro, la decisione sull’eliminazione del proprio profilo si limita alla mera indicazione della violazione di una specifica regola di un qualche social media. Per quanto la moderazione abbia potuto avere forma algoritmica (per l’individuazione e la valutazione dei tweet pubblicati dall’utente-presidente Trump durante l’insurrezione di Capitol Hill), ritengo che una particolare attenzione e moderazione ‘umana’ sia stata centrale. In termini generali, un’interferenza si realizza sulla fruizione delle funzionalità di espressione e interconnessione.

---

<sup>19</sup> S. Gorissen, *Weathering and Weaponizing the #TwitterPurge: Digital Content Moderation and the Dimension of Deplatforming*, in «Communication and Democracy», 58/1, 2014, p. 7.

Altri due livelli, meso e macro, sono individuabili. Il livello meso di *incremental deplatforming* si caratterizza per una dimensione cumulativa, tale per cui più account individuali vengono eliminati per la mancata conformità dei contenuti alle regole di un qualche social media. La prima citata fondamentale minaccia costituita dalla moderazione si inserisce qui. L'interferenza rende conto della rivendicazione di una censura lungo l'asse politico, "permette[ndo] di protestare contro i 'big Tech', [e] fungendo da grido di battaglia attorno al quale organizzarsi, rafforzare l'identificazione e fomentare il dissenso"<sup>20</sup>. Difatti, tale *deplatforming* cumulativo ha coinvolto quei soggetti-utenti accomunati dalla medesima prospettiva di senso politicizzata, ed è il caso dei sostenitori di Trump durante il suo primo mandato presidenziale (la c.d. *Twitter Purge*), e l'insurrezione di Capitol Hill. Segue il livello macro di *implosive deplatforming* che consiste in una interferenza strutturale di un intero social media, delle sue funzionalità e della sua utenza. Questo è il caso di Parler, piattaforma social non generalista che era divenuta rifugio per quegli utenti esclusi altrove. L'interferenza è stata realizzata dall'*Internet Service Provider* per la violazione dei suoi termini di servizio, a seguito degli eventi di Capitol Hill. La rivendicazione di una censura si è ripresentata e rafforzata.

Va, comunque, considerato che alcuni account oggetto di *deplatforming* sono stati reintegrati, e tra questi troviamo il caso di Trump. Anche il *deplatforming* di Parler è stato temporaneo, facendo sì che la piattaforma tornasse accessibile – nella sua homepage si legge "Parler. Real People. Unfiltered Voices. True Freedom"<sup>21</sup>. I toni emergenziali attorno la censura nella sua forma digitale sono, quindi, credibili? Nuovamente, occorre preoccuparsi? Gorissen presenta un elemento che ritengo interessante: la capacità adattiva di quella utenza che ha visto negate le possibilità di espressione e interconnessione. È stata la rivendicata censura a far sì che alcuni soggetti migrassero dai social media generalisti a quelli più di nicchia e permissivisti, e ciò potrebbe valere anche per coloro che non sono solo statunitensi. Se le questioni proprie del PC (la ricerca smodata di odio e discriminazione su base ascrivibile) interferiscono con un diritto di libertà di espressione entro i social media, sono esattamente queste a porre le condizioni per attaccare lo stesso PC. Ho parlato di un PC resistente nel tempo e che travalica i confini geografici, insinuandosi anche in spazi digitali. Ciò implicherebbe che il PC sia stato e rimanga tale come se nessuno l'avesse combattuto, o come se chiunque l'avesse subito

---

<sup>20</sup> *Ivi*, p. 9.

<sup>21</sup> <https://www.parler.com/>.

passivamente. Non nego che le accuse di -ismi e -fobie non sussistano, mi chiedo in quale misura la rivendicazione di una censura sia funzionale a rafforzare lo status di vittima, alzando la propria voce a causa dell'interferenza subita. Mi sembra, così, che il PC abbia avvantaggiato chi lo critica, e non solo chi dovrebbe proteggere – i gruppi minoritari su base ascrittiva nelle società di riferimento.

Forse, non si tratta solo di una questione di censura. Anche l'autocensura va analizzata, se il leggere tra le righe di alcuni è dissuasivo circa l'intenzione di esprimersi. Quanto una certa utenza fa e non fa può dipendere dall'altrui attenzione, implicando uno scenario di cautezza nella pubblicazione dei propri contenuti: evitare certe parole e questioni, nonostante si intende utilizzarle e affrontarle; tuttavia, da ciò non segue *ipso facto* che quei contenuti costituiscano una violazione delle regole dei social media, solo perché ci si esprime su specifiche caratteristiche ascrittive – direttamente o trasversalmente. All'attenzione (politicamente corretta) che alcuni soggetti mostrano costantemente su altri, si aggiunge la moderazione realizzabile dai social media. Quest'ultima fa sì che una certa utenza trovi delle ragioni per non pubblicare ciò che presume sia eliminabile. Ed anche se ci esprimesse, la moderazione algoritmica potrebbe essere interpretata quale inesatta nella sua valutazione, in relazione alla specifica intenzione di un soggetto tramite il suo contenuto e al contesto a cui rimanda. Per esempio, i contenuti d'odio costituiscono una categoria regolamentata, trattandosi di ciò che viene detto di e a soggetti alla luce delle loro caratteristiche ascrittive – definite protette dai social media. Come già detto, il leggere tra le righe proprio del PC rinvia anche ad una ricerca smodata di odio. Se questi sono tempi d'odio, chiunque odia? Un'incertezza su come esprimersi può emergere, con i soggetti che decidono di auto-censurarsi. Al riguardo, Simpson problematizza la questione dell'autocensura, muovendo dall'ipotesi secondo cui “i soggetti non sono uniformemente avversi al rischio nelle loro disposizioni comunicative”<sup>22</sup>; nella gestione dell'incertezza, alcuni eviteranno di esprimersi, altri non lo faranno.

L'avversione al rischio fa riferimento alla tendenza di un soggetto a preferire le opzioni con gli esiti più certi, rispetto a quelle con gli esiti meno certi. [...] Coloro che hanno un'alta avversione al rischio preferiscono *speech act* (o omissioni) con risultati

---

<sup>22</sup> R.M. Simpson, *Self-Censorship: The Chilling Effect and the Heating Effect*, in «Political Philosophy», 1/2, 2024, p. 354. Cfr. P. Cook – C. Heilmann, *Two Types of Self-Censorship: Public and Private*, in «Political Studies», 61, 2013.

prevedibili, mentre coloro che hanno una bassa avversione al rischio sono più indifferenti circa le conseguenze imprevedibili di ciò che esprimono<sup>23</sup>.

In una considerazione del posizionamento del PC lungo l'asse politico, se la *correttezza* è propria della radicale sinistra, una *scorrettezza* controversa è riconducibile alla destra che è vittima della limitazione di un diritto di libertà di espressione? Tale *scorrettezza* può essere interpretata tramite la questione della bassa avversione al rischio, e ipotizzo ciò alla luce della critica resistente al PC. Seguendo Simpson, se alcuni soggetti si auto-censurano, il discorso pubblico individuabile *anche* entro i social media dovrebbe essere caratterizzato da espressioni pacate e non provocatorie. È, invece, la presenza di soggetti con un basso livello di avversione al rischio (e quindi al rischio delle accuse di -ismi e -fobie e del loro stigma) a far sì che essi abbiano più spazio. Il discorso pubblico si intensifica, con la *scorrettezza* dei critici del PC che ritengo permetta una categorizzazione per opposizione rispetto a chi soccombe alla *correttezza*: si tratta di coloro che sfidano le interferenze poste dal PC, tramite un'audacia che manca ad altri e che prende la forma di valutazioni schiette e non accomodanti, ironia, ridicolizzazione, esaltazione del disaccordo e, in particolar modo, rifiuto di leggere tra le righe. Se da ciò muove la rivendicazione di una concettualizzazione di libertà di espressione, questa è alquanto specifica.

A questo punto, intendo riflettere sulla *correttezza* presentata e la *conformità* alle regole proprie dei social media. Affinché i contenuti siano pubblicabili e rimangano visibili, occorre che essi siano conformi a quelle regole che i social media definiscono. Sopra si è accennato ai contenuti d'odio regolamentati, e proprio questi costituiscono un aspetto peculiare della *content moderation*. Al giorno d'oggi, tutti i social media generalisti sono statunitensi, non considerando TikTok, costituendo un'eccezione al c.d. eccezionalismo statunitense, in termini della protezione costituzionale della libertà di espressione nella forma dello *hate speech* – a cui i contenuti d'odio sono rinviabili. I social media regolamentano, quindi, ciò che al governo è impedito fare. E forse, un'utenza statunitense è più sensibile, rispetto a qualunque altra, a tale incongruenza. La conformità alle regole potrebbe essere interpretata come un'interferenza da sé problematica. Se poi si individua un'intersezione con la

---

<sup>23</sup> *Ivi*, p. 355.

*correttezza* del PC, la questione si inspessisce. In tale scenario, preoccupazioni sono state presentate, fino a prendere forma in due normative statali, il cui scopo consiste nella limitazione della moderazione dei social media, facendo esplicitamente riferimento alla loro censura. Non solo gli utenti possono essere attenti su ciò viene espresso. Anche l'autorità politica intende esserlo, in un tentativo di regolamentare la relazione tra i soggetti politici-utenti e i social media<sup>24</sup>.

Nel testo della normativa del Texas<sup>25</sup> si legge, “[the] state has a fundamental interest in protecting the free exchange of ideas and information”; relativamente la normativa della Florida<sup>26</sup>, “the state has a substantial interest in protecting its residents from inconsistent and unfair actions by social media platforms”. In entrambi i casi, la *viewpoint* (la prospettiva di senso) dell'utenza e dei contenuti non giustifica la regolamentazione propria dei social media. La moderazione algoritmica, inoltre, non deve essere opaca, ponendo l'utenza nelle condizioni di venire a conoscenza di come essa viene utilizzata nella valutazione di qualunque contenuto. Considerando quest'ultimo aspetto, “[t]he disclosure [...] must be sufficient to enable users to make an informed choice regarding the purchase of or use of access to or services from the platform”, per la prima normativa; nella seconda, “[a] social media platform must publish the standards, including detailed definitions, it uses or has used for determining how to censor, deplatform, and shadow ban”.

La questione è ampia e tramite una selezione, ritengo che un elemento centrale riguardi una preoccupazione (dell'autorità politica e dei suoi soggetti politici che sono anche utenti) nei confronti di una libertà di espressione che risulta essere processata algebricamente. La riduzione della visibilità dei contenuti può essere qui analiticamente utile. Per Gillespie essa concerne “contenuti rischiosi, fuorvianti, salaci [ed] è oggi utilizzata da molte piattaforme per limitare qualsiasi contenuto che ritengono quasi in violazione delle proprie regole”<sup>27</sup>. A differenza del *deplatforming*, non si tratta di un'eliminazione, proprio perché i contenuti non vengono eliminati. I social media decidono ciò

---

<sup>24</sup> Si veda Robert Gorwa, *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*, Oxford University Press, New York 2024.

<sup>25</sup> Texas HB 20 (2021), <https://www.flsenate.gov/Session/Bill/2021/7072/>.

<sup>26</sup> Florida SB 7072 (2021), <https://capitol.texas.gov/BillLookup/Text.aspx?LegSess=872&Bill=HB20>.

<sup>27</sup> T. Gillespie, *Do Not Recommend? Reduction as a Form of Content Moderation*, in «Social Media+Society», 8/3, 2022, p. 1. Cfr. K. Cotter, “Shadowbanning is not a thing”: *black box gaslighting and the power to independently know and credibly critique algorithms*, in «Communication and Society», 26/6, 2021.

che deve essere meno visibile e a chi, nonostante si tratti di contenuti conformi. L'utenza non è, quindi, posta nelle condizioni di venire a conoscenza di qualunque contenuto pubblicato, anche se non credo che essa possa caratterizzarsi per un occhio onnipervasivo sui propri spazi digitali. Tale riduzione è stata ed è oggetto di dibattito anche entro la questione del PC. Alla luce del posizionamento lungo l'asse politico, i critici del PC si considerano vittime di un'invisibilità. Tuttavia, le rimostranze sono individuabili e non si può affermare con esattezza in che misura sarebbero più visibili, se la riduzione algoritmica non fosse realizzata. La moderazione rimane, ossia, nelle mani dei social media. Ritengo che lo svantaggio rivendicato possa essere utilizzato strategicamente, e non tanto perché alcuni decidono di lasciare gli spazi digitali generalisti, migrando in quelli meno ostili e più permissivisti. Anche rimanendovi il PC viene attaccato, in modo tale che esso sia catalizzatore di dissenso e di aggregazione in funzione oppositiva.

Quella che ho definito una libertà di espressione processata algoritmicamente rende conto, a mio avviso, di una valutazione algoritmica di quanto viene variamente espresso. Le regole di moderazione, relative ai contenuti pubblicabili o meno, ci sono sempre state. Queste sono state affinate, adattandosi ad errori, critiche e cambiamenti delle società in cui l'utenza usufruisce dei servizi dei social media. È da ricordare qui che la valutazione algoritmica (ed anche umana) realizzata non corrisponde alla valutazione di legittimità dell'esercizio del diritto di libertà di espressione. L'utenza è sì globale, ma la libertà di espressione rimane pur sempre un diritto riconosciuto, prima di tutto, ad un livello statale. Se la libertà di espressione è processata algoritmicamente, occorre che essa sia anche interattiva e, grazie a ciò, redditizia. Il dibattito del PC entro i social media genera flussi di dati? Credo si possa rispondere affermativamente. Daub analizza la *cancel culture* come fenomeno individuabile negli spazi digitali generalisti. Essa viene definita un *continuum* che muove dal PC e dal suo dibattito, nella misura in cui le due questioni si caratterizzano per la medesima preoccupazione. “Sia negli Stati Uniti sia all'estero, i discorsi sulla *cancel culture* propongono che stiamo vivendo una marea crescente di censura (di solito di sinistra)”<sup>28</sup>. “Gli eventi in una nicchia ricercata di Twitter, in un piccolo college o in una casa editrice [...], si trasformano improvvisamente

---

<sup>28</sup> A. Daub, *Cancel Culture Panic. How an American Obsession Went Global*, Stanford University Press, Stanford 2024, p. 14. Cfr. A. Dershowitz, *Cancel Culture. The Latest Attack on Free Speech*, Hot Books, New York, 2020; P. Norris, *Cancel Culture: Myth or Reality?*, « Political Studies », 71/1, 2023.

in grandi minacce per la società, la libertà e l'Occidente"<sup>29</sup>. Sembra, quindi, che le preoccupazioni possano essere costruite e adattate, generando flussi di dati e con lo spettro della censura che prende la forma della moderazione dei social media.

In una copertina del *New Yorker* del 1991, si legge “Are you politically correct? Am I guilty of racism, sexism, classism? Am I guilty of ageism, ableism, lookism? Am I logocentric? Do I say ‘Indian’ instead of ‘Native American’? ‘Pet’ instead of ‘animal companion’?”<sup>30</sup>. Niente di più attuale che le domande di Taylor. Da qui, se *i politicamente corretti* sono strenuamente attenti anche *i politicamente scorretti* lo sono.

---

<sup>29</sup> *Ivi*, pp. 5-6.

<sup>30</sup> J. Taylor, *Are you Politically Correct?*, *New Yorker*, 21 gennaio 1991.



Sesto San Giovanni (MI)  
via Monfalcone, 17/19



**AlboVersorio  
Edizioni**

& AlboVersorio Edizioni  
di Ass. NonsoloSophia  
nonsolosophia@gmail.com

© Metabasis.it, rivista semestrale di filosofia e comunicazione.  
Autorizzazione del Tribunale di Varese n. 893 del 23/02/2006.  
ISSN 1828-1567



Quest'opera è stata rilasciata sotto la licenza Creative Commons Attribuzione-NonCommerciale-NoOpereDerivate 2.5 Italy. Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by-nc-nd/2.5/it/> o spedisci una lettera a Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.