Chi o cosa controlla ciò che non vediamo sul web?

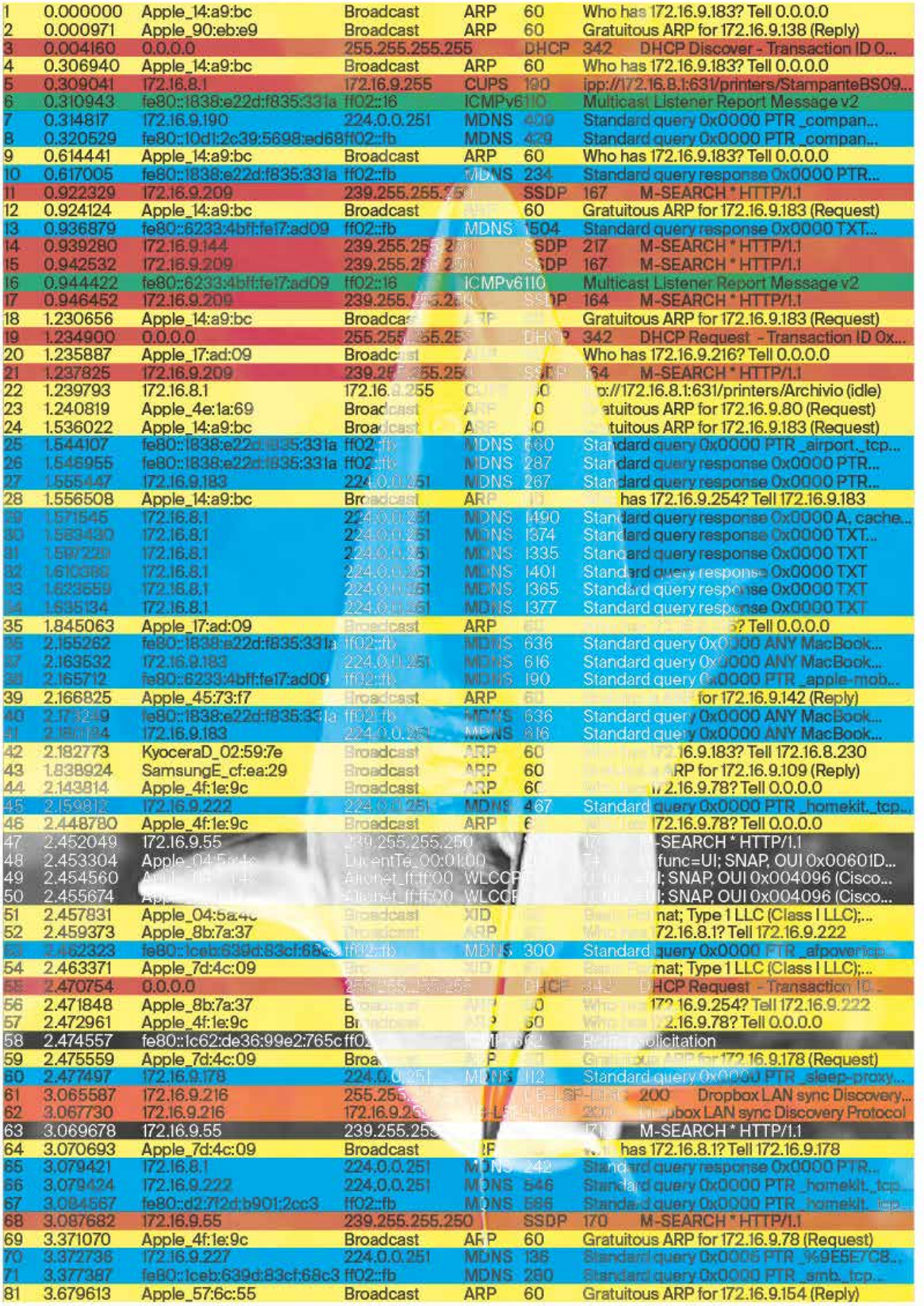
Eleonora Benecchi, docente presso l'Università della Svizzera italiana

L'ultima cosa che pensiamo quando ci troviamo a riflettere sui contenuti che circolano, apparentemente incontrollati, sul Web e soprattutto sui social media è quello che non vediamo. Rendiamo questo concetto più concreto. Perché quando apriamo la nostra pagina Facebook, se ne abbiamo una, non vediamo come prima cosa delle immagini pornografiche o violente? Viviamo in un'epoca in cui l'illusione che le persone non postino online foto di nudo, video erotici, immagini di violenza inflitta o subita, è svanita da tempo. Sappiamo che le persone pubblicano online questo tipo di materiale, le ricerche ci dicono che molti lo fanno anche tramite social media e, dunque, come mai questo tipo di immagini non invadono le nostre bacheche? Perché esiste una costante moderazione dei contenuti che sono pubblicati online. Ed ecco allora una questione direttamente collegata, ma rimossa dal pensiero comune: chi o cosa controlla cosa vediamo, ma soprattutto cosa non vediamo sul Web? Molti sono convinti che si tratti di un controllo automatizzato affidato ai famigerati algoritmi di cui così tanto, ma con così poca chiarezza, si sente parlare. Ogni piattaforma social li utilizza per mettere ordine nel flusso costante e confuso di contenuti pubblicati online, quindi questi algoritmi saranno probabilmente utilizzati anche per garantire la 'sicurezza' degli utenti, evitare di esporli a immagini indesiderate o che violano le regole del sito a cui sono iscritti.

Ci sono voluti due documentari, *The moderators* di Ciaran Cassidy e Adrian Chen, e *The Cleaners* di Hans Block e Moritz Riesewieck, per portare alla luce il lato umano che nutre l'industria sommersa della moderazione online. Ignora. Cancella. Ignora. Cancella. Sono le parole che sentiamo più spesso nei due documentari in questione. Sono anche i due confini estremi dello spazio in cui si muovono i cosiddetti 'moderatori' o 'pulitori': le persone che ripuliscono la Rete dalla 'sporcizia', ovvero da tutti i contenuti che violano le regole di condotta delle piattaforme online che li hanno assunti. Stiamo parlando di individui che lavorano per alcune delle aziende più popolari al mondo, da Facebook, a Youtube, passando per Instagram, ma che non possono parlare del loro lavoro, né di chi li ha impiegati per svolgerlo. Quello che scopriamo guardando questi documentari, ma anche leggendo le inchieste che negli ultimi anni sono state pubblicate sulla moderazione dei contenuti online, è che esiste una vera e propria economia ombra sviluppata attorno alla scintillante e ampiamente promossa industria dei social media. Eppure, come sostenuto dalla scrittrice e fotografa americana Susan Sontag, riflettere su quello che i media (di cui i social media e il Web più in generale fanno parte) mostrano o nascondono, su quello che si decide di pubblicare o di non pubblicare, è fondamentale per il progresso culturale e sociale. Perché le immagini non possono creare una posizione morale, ma possono rafforzarla: "Immagini come quella che nel 1972 comparve sulle prime pagine di quasi tutti i quotidiani del mondo – la bambina sud vietnamita che, irrorata dal napalm americano, correva su una strada verso l'obiettivo, a braccia aperte e urlando di dolore – contribuirono probabilmente ad accrescere l'avversione dell'opinione pubblica alla guerra" scrive la Sontag (*Sulla Fotografia*, 1977). Si tratta di una riflessione complessa e la stessa Sontag nel corso della sua vita cambierà radicalmente idea. Negli anni Settanta afferma che il sovraffollamento di immagini violente, come le foto di guerra, rischia di anestetizzare e paralizzare chi le guarda, mentre alla fine della sua vita assume una posizione contraria: nel suo ultimo saggio *Davanti al dolore degli altri* (2003) invita a "lasciarsi ossessionare dalle immagini più atroci […]. Quelle immagini dicono: ecco ciò che gli esseri umani sono capaci di fare, ciò che – entusiasti e convinti d'essere nel giusto – possono prestarsi a fare. Non dimentichiamolo". Al di là di quale posizione decidiamo di condividere è chiaro che il lavoro di chi ogni giorno è chiamato a scegliere cosa deve essere visto e cosa deve essere nascosto sul Web non può e non deve essere mantenuto segreto. Dovrebbe invece essere oggetto di dibattito pubblico, un dibattito che è politico, ma anche etico e culturale. Questo, però, non sarà possibile fino a che non avremo informazioni concrete su cosa questo tipo di moderazione online comporta, su come le persone che svolgono questo lavoro sono effettivamente formate, da chi dipendono, quali influenze culturali e politiche subiscono.

Iniziamo dunque la nostra riflessione sul tema della moderazione online da ciò che sappiamo, grazie alle informazioni fornite dai già citati documentari di inchiesta e da inchieste giornalistiche vere e proprie, tra cui *Inside Facebook* di Hannes Grassegger e Till Krause.

Cosa fanno, in pratica, questi moderatori della Rete? Rispondono alle segnalazioni degli utenti (definite in gergo 'tickets') valutando se la segnalazione è legittima e decidendo poi se il video o l'immagine accusata di violare i codici di condotta del social media può rima-



1	0.000000	Apple_14:a9:bc	Broadcast	ARP	60	Who has 172.16.9.183? Tell 0.0.0.0
2	0.000971	Apple_90:eb:e9	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.138 (Reply)
3	0.004160	0.0.0.0	255.255.255.255		DHCP	342 DHCP Discover - Transaction ID 0...
4	0.306940	Apple_14:a9:bc	Broadcast	ARP	60	Who has 172.16.9.183? Tell 0.0.0.0
5	0.309041	172.16.8.1	172.16.9.255	CUPS	190	ipp://172.16.8.1:631/printers/StampanteBS09...
6	0.310943	fe80::1838:e22d:f835:331a	ff02::16	ICMPv6	110	Multicast Listener Report Message v2
7	0.314817	172.16.9.190	224.0.0.251	MDNS	409	Standard query 0x0000 PTR _compan...
8	0.320529	fe80::10d1:2c39:5698:ed68	ff02::fb	MDNS	429	Standard query 0x0000 PTR _compan...
9	0.614441	Apple_14:a9:bc	Broadcast	ARP	60	Who has 172.16.9.183? Tell 0.0.0.0
10	0.617005	fe80::1838:e22d:f835:331a	ff02::fb	MDNS	234	Standard query response 0x0000 PTR...
11	0.922329	172.16.9.209	239.255.255.250	SSDP	167	M-SEARCH * HTTP/1.1
12	0.924124	Apple_14:a9:bc	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.183 (Request)
13	0.936879	fe80::6233:4bff:fe17:ad09	ff02::fb	MDNS	1504	Standard query response 0x0000 TXT...
14	0.939280	172.16.9.144	239.255.255.250	SSDP	217	M-SEARCH * HTTP/1.1
15	0.942532	172.16.9.209	239.255.255.250	SSDP	167	M-SEARCH * HTTP/1.1
16	0.944422	fe80::6233:4bff:fe17:ad09	ff02::16	ICMPv6	110	Multicast Listener Report Message v2
17	0.946452	172.16.9.209	239.255.255.250	SSDP	164	M-SEARCH * HTTP/1.1
18	1.230656	Apple_14:a9:bc	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.183 (Request)
19	1.234900	0.0.0.0	255.255.255.255	DHCP	342	DHCP Request - Transaction ID 0x...
20	1.235887	Apple_17:ad:09	Broadcast	ARP	60	Who has 172.16.9.216? Tell 0.0.0.0
21	1.237825	172.16.9.209	239.255.255.250	SSDP	164	M-SEARCH * HTTP/1.1
22	1.239793	172.16.8.1	172.16.9.255	CUPS	60	ipp://172.16.8.1:631/printers/Archivio (idle)
23	1.240819	Apple_4e:1a:69	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.80 (Request)
24	1.536022	Apple_14:a9:bc	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.183 (Request)
25	1.544107	fe80::1838:e22d:f835:331a	ff02::fb	MDNS	660	Standard query 0x0000 PTR _airport._tcp...
26	1.546955	fe80::1838:e22d:f835:331a	ff02::fb	MDNS	287	Standard query response 0x0000 PTR...
27	1.555447	172.16.9.183	224.0.0.251	MDNS	267	Standard query response 0x0000 PTR...
28	1.556508	Apple_14:a9:bc	Broadcast	ARP	60	Who has 172.16.9.254? Tell 172.16.9.183
29	1.571545	172.16.8.1	224.0.0.251	MDNS	1490	Standard query response 0x0000 A, cache...
30	1.583430	172.16.8.1	224.0.0.251	MDNS	1374	Standard query response 0x0000 TXT...
31	1.597229	172.16.8.1	224.0.0.251	MDNS	1335	Standard query response 0x0000 TXT
32	1.610389	172.16.8.1	224.0.0.251	MDNS	1401	Standard query response 0x0000 TXT
33	1.623559	172.16.8.1	224.0.0.251	MDNS	1365	Standard query response 0x0000 TXT
34	1.635134	172.16.8.1	224.0.0.251	MDNS	1377	Standard query response 0x0000 TXT
35	1.845063	Apple_17:ad:09	Broadcast	ARP	60	Who has 172.16.9.1? Tell 0.0.0.0
36	2.155262	fe80::1838:e22d:f835:331a	ff02::fb	MDNS	636	Standard query 0x0000 ANY MacBook...
37	2.163532	172.16.9.183	224.0.0.251	MDNS	616	Standard query 0x0000 ANY MacBook...
38	2.165712	fe80::6233:4bff:fe17:ad09	ff02::fb	MDNS	190	Standard query 0x0000 PTR _apple-mob...
39	2.166825	Apple_45:73:f7	Broadcast	ARP	60	Who has 172.16.9.142 (Reply)
40	2.173249	fe80::1838:e22d:f835:331a	ff02::fb	MDNS	636	Standard query 0x0000 ANY MacBook...
41	2.180184	172.16.9.183	224.0.0.251	MDNS	616	Standard query 0x0000 ANY MacBook...
42	2.182773	KyoceraD_02:59:7e	Broadcast	ARP	60	Who has 172.16.9.183? Tell 172.16.8.230
43	1.838924	SamsungE_cf:ea:29	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.109 (Reply)
44	2.143814	Apple_4f:1e:9c	Broadcast	ARP	60	Who has 172.16.9.78? Tell 0.0.0.0
45	2.159812	172.16.9.222	224.0.0.251	MDNS	467	Standard query 0x0000 PTR _homekit._tcp...
46	2.448780	Apple_4f:1e:9c	Broadcast	ARP	60	Who has 172.16.9.78? Tell 0.0.0.0
47	2.452049	172.16.9.55	239.255.255.250	SSDP	170	M-SEARCH * HTTP/1.1
48	2.453304	Apple_04:5a:4c	LucentTe_00:01:00		74	Ethernet II, func=UI; SNAP, OUI 0x00601D...
49	2.454560	Apple_4f:1e:9c	Xircom_ff:ff:00	WLCCP	60	Ethernet II; SNAP, OUI 0x004096 (Cisco...
50	2.455674	Apple_4f:1e:9c	Xircom_ff:ff:00	WLCCP	60	Ethernet II; SNAP, OUI 0x004096 (Cisco...
51	2.457831	Apple_04:5a:4c	Broadcast	XID	60	Ethernet; Format; Type 1 LLC (Class I LLC);...
52	2.459373	Apple_8b:7a:37	Broadcast	ARP	60	Who has 172.16.8.1? Tell 172.16.9.222
53	2.462323	fe80::1ceb:639d:83cf:68c3	ff02::fb	MDNS	300	Standard query 0x0000 PTR _afpovertcp...
54	2.463371	Apple_7d:4c:09	Broadcast	XID	60	Ethernet; Format; Type 1 LLC (Class I LLC);...
55	2.470754	0.0.0.0	255.255.255.255	DHCP	342	DHCP Request - Transaction ID...
56	2.471848	Apple_8b:7a:37	Broadcast	ARP	60	Who has 172.16.9.254? Tell 172.16.9.222
57	2.472961	Apple_4f:1e:9c	Broadcast	ARP	60	Who has 172.16.9.78? Tell 0.0.0.0
58	2.474557	fe80::1c62:de36:99e2:765c	ff02::2	ICMPv6	62	Router Solicitation
59	2.475559	Apple_7d:4c:09	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.178 (Request)
60	2.477497	172.16.9.178	224.0.0.251	MDNS	112	Standard query 0x0000 PTR _sleep-proxy...
61	3.065587	172.16.9.216	255.255.255.255	DB-LSP-DISC	200	Dropbox LAN sync Discovery...
62	3.067730	172.16.9.216	172.16.9.255	DB-LSP	200	Dropbox LAN sync Discovery Protocol
63	3.069678	172.16.9.55	239.255.255.250	SSDP	170	M-SEARCH * HTTP/1.1
64	3.070693	Apple_7d:4c:09	Broadcast	ARP	60	Who has 172.16.8.1? Tell 172.16.9.178
65	3.079421	172.16.8.1	224.0.0.251	MDNS	242	Standard query response 0x0000 PTR...
66	3.079424	172.16.9.222	224.0.0.251	MDNS	546	Standard query 0x0000 PTR _homekit._tcp...
67	3.084567	fe80::d2:712d:b901:2cc3	ff02::fb	MDNS	586	Standard query 0x0000 PTR _homekit._tcp...
68	3.087682	172.16.9.55	239.255.255.250	SSDP	170	M-SEARCH * HTTP/1.1
69	3.371070	Apple_4f:1e:9c	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.78 (Request)
70	3.372736	172.16.9.227	224.0.0.251	MDNS	136	Standard query 0x0000 PTR _9B9E5E7CB...
71	3.377387	fe80::1ceb:639d:83cf:68c3	ff02::fb	MDNS	280	Standard query 0x0000 PTR _smb._tcp...
81	3.679613	Apple_57:6c:55	Broadcast	ARP	60	Gratuitous ARP for 172.16.9.154 (Reply)

Ruben Campoy Bähler
4° anno di grafica – CSIA

nere online o deve essere rimossa. Ecco che già si profila il primo problema legato a questo tipo di sistema: il regolamento di social come Facebook è *user based regulation*: i contenuti non sono regolamentati prima della pubblicazione, ma solo dopo. Dunque le immagini o i video 'incriminati' sono già online e visibili a tutti, prima che vengano eventualmente rimossi. In questo contesto, non stupisce che a maggio 2018 il video di un omicidio sia rimasto online su Facebook per due ore. La vicenda, aspramente criticata dalla politica e dai media, ha spinto il CEO Mark Zuckerberg a richiedere l'assunzione di altri 3'000 moderatori che andranno ad aggiungersi al cosiddetto *Community Operations Team* nel tentativo di "rendere la comunità di Facebook più sicura per tutti". C'è evidentemente un problema di numeri in questo sistema. Facebook afferma di avere 7'500 moderatori nel mondo, di cui l'200 solo in Germania, e pare che abbia previsto di assumerne altri 3'000 entro il 2019. Tuttavia la mole di lavoro che queste persone si trovano a dovere gestire è impari. Secondo i dati pubblicati dalle principali inchieste svolte sul tema, tra cui la già citata *Inside Facebook*, sono più di 6 milioni i 'ticket' generati ogni settimana. I moderatori intervistati dai cineasti tedeschi di *The Cleaners* guadagnano da 1 a 3 dollari l'ora per moderare una media di 8'000 ticket al giorno. Una realtà che si ripete in altre parti del mondo. Perché l'industria ombra della pulizia digitale non opera solo in Paesi in via di sviluppo, ma anche nella moderna Germania, dove a Berlino ed Essen si trovano due dei più grandi centri di moderazione di Facebook al mondo, come documentato dall'inchiesta del giornale tedesco Süddeutsche Zeitung (SZ). Perché proprio la Germania? Paradossalmente perché in questo Paese le leggi sull'hate speech e sulla moderazione dei contenuti online sono particolarmente rigide. Da gennaio 2017, il Netzwerkdurchsetzungsgesetz (NetzDG) obbliga Facebook, Twitter e YouTube a rimuovere hate speech, fake news, e materiali illegali entro 24 ore di tempo o pagare multe fino a 60 milioni di dollari. Questo impone alle piattaforme online di operare una costante e rigida moderazione dei contenuti pubblicati per cui sono ritenuti responsabili. Di fatto questo tipo di legislazione mette Facebook e compagni allo stesso livello della stampa, assegnando a queste piattaforme un ruolo editoriale e dunque imponendo un certo livello di responsabilità sui contenuti pubblicati. Il risultato è che i moderatori sembrano non essere mai abbastanza. Ma questo è anche dovuto al fatto, come abbiamo visto, che la mole di lavoro richiesta a ogni moderatore esige ritmi di gestione disumani. La questione è reale e stringente. Nell'ultimo anno video di suicidi, omicidi, sesso esplicito, violenza inflitta o subita sono stati diffusi con una frequenza allarmante sul social media più popolare in occidente e con oltre 2 miliardi di iscritti al mondo.

Questo ci porta al secondo problema collegato alla scelta di affidare al controllo umano la moderazione online. Il ruolo di moderatore è spesso presentato come un lavoro prettamente tecnico: ci sono dei protocolli e delle linee guida da seguire per decidere quali segnalazioni accettare e quali rifiutare. Nel documentario *The Moderators*, una delle poche fonti disponibili che ci mostra (con occhio critico) il processo di formazione attraverso cui i neo-assunti devono passare per diventare moderatori, vediamo che il lavoro di moderazione viene presentato come qualcosa di neutrale: a stimolo risposta. Vediamo anche che le persone assunte per questo tipo di lavoro non hanno idea di cosa andranno concretamente a fare e nessuna esperienza in materia di moderazione di contenuti online o offline. Eppure viene chiesto loro di "essere estremamente bravi a giudicare le cose, capirle fino in fondo, perché non sono concessi errori". Nessun errore è concesso a persone che spesso sono al loro primo impiego e non hanno alcuna esperienza in materia. Una visione quanto mai idealistica. Meno idealistico è il lavoro effettivo che i moderatori andranno a svolgere. Sebbene sia presentato come un lavoro simile a chi assegna i ratings ai film hollywoodiani, categorizzare i diversi contenuti e bloccare quello che potrebbe essere lesivo della sensibilità dell'utente comune, di fatto implica guardare immagini o video di abusi, spesso su minori, torture, omicidi, suicidi, scenari di guerra, decapitazioni, ogni giorno più ore al giorno. Ecco allora che se scendiamo nel concreto delle storie personali dei moderatori capiamo che una giornata tipo consiste nel guardare fino a 8'000 immagini o video di autolesionismo, spesso senza ricevere alcun supporto psicologico e rimanendo di fatto invisibili al mondo. Nelle Filippine il moderatore che si è trovato in questa situazione, più volte denunciata ai suoi capi diretti come 'ingestibile', ha finito per suicidarsi. Ma perché manca il supporto psicologico o quello che viene offerto non è sufficiente? Perché le denunce interne sono inascoltate o non vengono formalmente presentate? Il motivo è strutturale: come denunciato da varie inchieste aperte contro il sistema di moderazione di giganti come Facebook e Youtube,

le persone responsabili della moderazione spesso sono formate in maniera superficiale per il lavoro che sono chiamate a svolgere e nella quasi totalità dei casi non sono assunte direttamente dalle aziende, ma da compagnie terze, in outsourcing: i cosiddetti 'clienti', che sono entità astratte e innominabili, a meno che non si voglia rischiare una denuncia per violazione degli obblighi contrattuali. Perché tutta questa segretezza? Perché le grandi aziende che dominano il mercato del Web non vogliono pubblicizzare il fatto che i contenuti che ognuno di noi individualmente posta su profili spesso privati sono visti da altri esseri umani chiamati a giudicarne l'appropriatezza: i moderatori hanno accesso a un enorme numero di dati sensibili relativi agli utenti che pubblicano contenuti. In questo senso l'outsourcing è una pratica che deresponsabilizza l'azienda principale nel caso dovessero sorgere problemi e che lascia i lavoratori senza garanzie o supporto. La segretezza ha però anche a che fare con un secondo aspetto, relativo all'impatto che il lavoro di moderazione ha su chi lo svolge. I contenuti pubblicati online dagli utenti sono un prodotto preziosissimo, che tuttavia, come molti prodotti di grande valore, genera anche dei sotto-prodotti indesiderati, delle scorie, se vogliamo. Queste scorie devono essere gestite. La vera domanda, però, è 'come'; e la risposta non è scontata. Al momento questi sotto-prodotti indesiderati vengono gestiti in outsourcing, ovvero scaricati su lavoratori precari e Paesi in via di sviluppo. Le persone adibite a trattare e smaltire le nostre scorie non sono adeguatamente protette e tutelate. Non si tratta solo di abuso del precariato, ma anche di un vero e proprio abuso mentale: le condizioni lavorative dei moderatori di contenuti online hanno effetti devastanti anche sul piano psicologico, come documentato dalle inchieste del Wall Street Journal (*The Worst Job in Technology: Staring at Human Depravity to Keep it off Facebook*, 27 Dicembre 2017) e comprovato dalle sempre più frequenti denunce nei confronti delle compagnie di Silicon Valley da parte di ex-moderatori che dichiarano di avere subito irreparabili danni fisici e psicologici a causa del lavoro che sono stati chiamati a svolgere. Nel solo 2017 già tre denunce sono finite in tribunale: due contro Microsoft e una contro Facebook. Poche se si pensa al numero di moderatori attivi al mondo, che si aggirano intorno ai 150'000 secondo dati raccolti da Al Jazeera nel 2017, ma molte se consideriamo l'enorme difficoltà di portare avanti denunce di questo tipo quando si è assunti in outsourcing per fare un lavoro di cui nessuno sa nulla.

Eppure questo tipo di lavoro continua ad attirare molte persone. Non solo per motivazioni economiche, ma anche per ragioni sociali e morali. "Noi siamo l'ultimo baluardo della sicurezza online", "senza di noi la Rete sarebbe un luogo meno sicuro", "le nostre scelte possono salvare o condannare le persone", "senza regole ci sarebbe il caos, dobbiamo avere occhi ovunque", sono solo alcune delle frasi che i moderatori intervistati in *The Cleaners* e *The Moderators* usano per giustificare il loro attaccamento a un lavoro che appare quanto meno brutale a un occhio esterno.

Poniamo dunque il caso che il lavoro dei moderatori online fosse 'messo in sicurezza', e che le aziende che dominano la socializzazione online prendessero davvero su di sé la responsabilità della moderazione, informando pienamente riguardo alle richieste e alle caratteristiche del lavoro che offrono, formando adeguatamente il personale e fornendogli il supporto psicologico necessario. Problema risolto? Evidentemente no, perché le questioni sollevate dalla moderazione dei contenuti online non sono solo di natura individuale, ovvero non riguardano solo le condizioni lavorative e gli eventuali danni, psicologici e fisici, che i 'pulitori della Rete' possono subire nel corso della loro opera.

Una delle domande al cuore di tutte le ricerche sul ruolo dei moderatori della Rete ha una natura più generale e attiene alla sfera culturale e politica, oltre che etica: i cosiddetti 'moderatori' agiscono semplicemente come un servizio di pubblica utilità che difende la sicurezza e il benessere degli utenti, e dunque in maniera neutrale, o le loro decisioni sono di fatto decisioni 'editoriali' e dunque hanno un impatto pesante sulla libertà di espressione in Rete?

Fino a che punto è legittimo rimuovere le immagini di un bombardamento su civili, perché oggettivamente violente, quando la documentazione di tale violenza può rappresentare un atto politico contro la dittatura responsabile di tale atto?

Siamo d'accordo che un'immagine di Donald Trump ritratto nudo e 'debole' da parte di un'artista impegnata in una critica politica debba essere rimossa perché offende la virilità del personaggio pubblico? O questo è forse un atto di censura determinato anche dalle origini culturali e dalle credenze religiose di una moderatrice che non è stata formata per svolgere un ruolo editoriale, ma solo per eseguire un compito meccanico?

Qual è la differenza tra hate speech e free speech in questo ambito?
Si tratta di domande che precedono e vanno ben al di là dell'arrivo e dello sviluppo dei social media. Il problema della moderazione online, infatti, non nasce con loro. Questo tipo di discussioni esisteva già ai tempi dei forum, ovvero le comunità di discussione online, quando ancora Facebook, Yotube, Instagram e Twitter non erano neanche dei progetti in fase di sviluppo. La differenza è che a quei tempi veniva ancora attuata una moderazione a priori: i commenti, le immagini, i video postati, venivano sottoposti a controllo prima di essere pubblicati. Il ritmo di pubblicazione era più lento, così come maggiore era la nostra capacità di attesa. Chi si ricorda l'inconfondibile suono del modem a 56k che si connette alla Rete per dare al computer di casa accesso a Internet, sa bene quanta pazienza la Rete chiedesse ai suoi utenti. Oggi, dato il numero di contenuti pubblicati sui social ogni minuto, affidare a degli umani questo tipo di moderazione a priori sarebbe impensabile: i ritardi nella pubblicazione di commenti e risposte non sarebbero considerati accettabili. I news media applicano ancora la moderazione a priori, ma è evidente che in questo caso i contenuti pubblicati sono controllati 'dall'alto' e la moderazione riguarda i commenti dei lettori, non in generale i contenuti pubblicati online dagli utenti. Inoltre la moderazione a priori in questo caso è necessaria, dato che i media di informazione sono responsabili non solo dell'articolo che pubblicano, ma anche dei commenti che appaiono sulle loro pagine, trattandosi sempre di contenuti 'editoriali'. Ecco perché leggi come quella tedesca sono così importanti, dato che richiamano Facebook e i social media alla loro responsabilità editoriale. Da parte loro invece i social stanno tentando tutte le strade possibili per rifiutare di mettersi in una posizione di questo tipo, perché accettare un ruolo editoriale significherebbe assumere in prima persona e con piena responsabilità il compito della moderazione dei contenuti pubblicati dagli utenti.
Quella che viene offerta al momento è una moderazione 'umana', e dunque imperfetta, svolta a posteriori e senza che sia chiaro chi dovrebbe prendersi la responsabilità delle scelte attuate. Non sono responsabili i social, che assumono in outsourcing, né le compagnie terze, che sono solo dei tramiti, ma neppure possiamo pensare che la responsabilità possa essere individuale. Perché a ben guardare le linee guida e i modelli decisionali a cui i moderatori fanno riferimento sono imposti dall'alto. Il problema è chi c'è 'in alto'? Chi decide cosa si deve eliminare e cosa invece deve rimanere online?
Alcuni suggeriscono una soluzione tecnologica: se i moderatori umani non sono sufficienti a garantire la sicurezza degli utenti sul Web e se, d'altra parte, il lavoro di moderazione ha un impatto così negativo su chi lo svolge, si dovrebbe investire su dei moderatori 'robotici'. Il problema sta proprio nella parola 'investire'. Allo stato attuale l'intelligenza artificiale non è ancora in grado di interpretare il contesto nel quale i contenuti sono pubblicati o a cui sono legati e neppure le cosiddette zone grigie, collegate ad esempio all'uso di ironia o delle forme satiriche. L'iconica foto della bambina che corre nuda per salvarsi dagli orrori della guerra in Vietnam sarebbe inevitabilmente eliminata da un filtro tecnologico programmato per eliminare il connubio nudità/minori. La moderazione automatica dei contenuti esiste, ma è all'inizio del suo sviluppo: i software messi a punto dalle aziende leader del settore, Twitter e Youtube, sono ancora poco sofisticati e incapaci di prendere decisioni complesse. Alcuni social applicano delle soluzioni intermedie, ma verrebbe da dire che sono solo soluzioni di comodo: in America Google fa segnalare i contenuti violenti o di hate speech come di bassa qualità in modo che finiscano in fondo alle classifiche di ranking.
Al momento è la moderazione umana in grado di ripulire le nostre bacheche dai contenuti indesiderati e di proteggere la nostra esperienza online dagli orrori pubblicati in Rete. Tuttavia dobbiamo trovare un modo di tutelare i moderatori, perché questo lavoro, necessario, non si trasformi in una forma di abuso su chi lo pratica. Il primo passo è senza dubbio portare questo problema alla luce, farlo emergere come oggetto di discussione pubblica. Cominciamo con il rompere il velo di segretezza che avvolge la circolazione e la moderazione dei contenuti online. Diamo un volto e una storia alle persone a cui è affidato il compito di garantire la nostra sicurezza online. Chiediamoci se stiamo affidando questo compito alle persone giuste, se le stiamo formando nel modo giusto, se stiamo chiedendo loro un impegno equo. Bisogna pretendere trasparenza rispetto a tutto il processo di moderazione dei contenuti online, a partire dalla formulazione delle linee guida che i moderatori, ultimo anello di questa catena, sono tenuti a seguire. Più rivolgiamo la nostra attenzione a questo tema, più le aziende del Web dovranno rendere

Luca Riva
4° anno di grafica – CSIA

conto, pubblicamente, di come gestiscono e controllano quello che vediamo, e soprattutto quello che non vediamo, online. Non solo, come accade adesso, dopo uno scandalo o una denuncia arrivata sulle prime pagine dei giornali. Moderazione umana o tecnologica? Controllo privato o pubblico della circolazione dei contenuti? Qualunque soluzione decideremo di implementare e su qualunque strada decideremo di investire, fondamentale è che sia il risultato di una discussione pubblica e informata che ci coinvolga in quanto cittadini prima ancora che utenti della Rete.

La scuola è senza dubbio uno dei luoghi centrali in cui aprire e sviluppare il dibattito su questa realtà poco conosciuta: i giovani, infatti, sono al tempo stesso gli utenti più assidui della Rete e i cittadini del futuro. Per questo occorre stimolarli a sviluppare spirito critico nei confronti delle grandi aziende del web anche informandoli su come i contenuti del Web vengono filtrati e controllati, soprattutto quando queste pratiche comportano la violazione di diritti umani.