

# MOĆ BEZ ODGOVORNOSTI: POLITIKE MODERACIJE SADRŽAJA *ONLINE*-PLATFORMI U BORBI PROTIV INFODEMIJE

Iva Nenadić

Fakultet političkih znanosti  
Sveučilište u Zagreb  
E-mail: iva.nenadic@fpzg.hr

DOI: 10.20901/an.17.04  
Pregledni rad  
Prihvaćeno: studeni 2020.

**Sažetak** Rad sadržava pregled politika vodećih online-platforni u borbi protiv dezinformacija i infodemije kako bi se promislile njihove moguće izravne ili neizravne, kratkoročne ili dugoročne posljedice za medijski pluralizam i informirano građanstvo, uzimajući u obzir značenje platformi u oblikovanju informacijskog okruženja. Metoda istraživanja je analiza dokumenata koje su odabrane platforme objavile od početka koronakrize u veljači do 31. srpnja 2020. kako bi komunicirale promjene svojih politika moderacije sadržaja kao posljedicu te krize. Uključena su i izvješća, koja su u kolovozu 2020. dostavile Europskoj komisiji, o mjerama koje su primijenile kako bi ograničile širenje dezinformacija povezanih s bolešću COVID-19. Analiza sugerira da su platforme nedvojbeno poduzele niz potencijalno korisnih mjera kako bi odgovorile na izazove širenja lažnih, obmanjujućih i štetnih informacija u pandemiji, ali gotovo svaka od tih mjera sadržava i potencijalan rizik za slobodu izražavanja i medijski pluralizam. Ti rizici ne proizlaze toliko iz samih inicijativa koliko iz nedostatka regulatornog okvira i neravnoteže između odgovornosti i moći koju platforme posjeduju.

**Ključne riječi** dezinformacije, infodemija, COVID-19, online-platforme, politike moderacije sadržaja, medijski pluralizam

## Uvod

I prije no što je proglasila pandemiju bolesti uzrokovane novim koronavirusom, Svjetska zdravstvena organizacija (SZO) u veljači 2020. glasno je ukazivala na problem "infodemije", odnosno prevelika broja informacija koje cirkuliraju javnim prostorom, pri čemu su neke pogrešne ili namjerno manipulativne (dezinformacije). Infodemija se pojavljuje u događajima visokog intenziteta kada se u kratkom razdoblju i zbog velikog interesa jav-

nosti jako povećava broj informacija među kojima je teško, na prvi pogled, razlikovati glasine i dezinformacije od vjerodostojnih vijesti (PAHO 2020). Problem širenja pogrešnih i namjerno manipuliranih informacija jedan je od ključnih izazova domene *online*-platforni već neko vrijeme, ali u kontekstu infodemije predstavlja dodatni izazov za javno zdravlje. Prevelik broj informacija ljudima ponekad otežava pronalazak pouzdanih izvora, obavijesti i nužnih uputa, a pogrešne informacije mogu izravno ugroziti živote.

*Online*-platforme, poput društvenih medija, tražilica i agregatora vijesti, postaju sve većem broju ljudi primaran izvor informiranja (Standardni Eurobarometar 92, 2019). Na taj način privatne, uglavnom američke korporacije, vođene neoliberalnim tržišnim vrijednostima, postaju ključni akteri javne informacijske sfere i demokratskih sustava diljem svijeta (Van Dijck 2020). Iako same ne proizvode sadržaj, platforme objedinjuju ponudu različitih medija i drugih izvora te ih korisnicima nude kroz algoritamski proces personalizacije utemeljen na različitim čimbenicima, uključujući prethodne preferencije korisnika i njihove druge (osobne) podatke. Takvi algoritamski preporučitelji sadržaja smatraju se nužnima kako bi korisnicima omogućili snalaženje u golemom broju informacija dostupnih *online* (Helberger 2009; Gauch i dr. 2007; Oulasvirta i Blom 2007). No sve veći broj autora ih vidi i kao potencijalnu prijetnju pluralizmu i demokraciji, kako zbog toga što njihovo djelovanje nije transparentno (Zuiderveen Borgesius i dr. 2016) tako i zbog rizika da kroz personaliziran odabir sadržaja fragmentiraju javnu sferu, uskraćuju korisnicima drugačije poglede i mišljenja, a time i priliku za demokratski dijalog (Sunstein 2002, 2009; Pariser 2011).

Uz to što sadržaj preporučuju, platforme ga prema određenim kriterijima i moderiraju (filtriraju i uklanjaju) te kuriraju (ističući određeni sadržaj) kako bi zadovoljile ukus pojedinačnog korisnika i osigurale njegov duži boravak na platformi i veći angažman, što povećava prihod od oglašavanja (Gillespie 2018). Poslovni model *online*-platformi temelji se, dakle, na što većem angažmanu korisnika i zadržavanju njihove pozornosti, koja se potom prodaje oglašivačima. Takav poslovni model prioritzira sadržaje koji ostvaruju veći angažman i potencijal viralnosti, a upravo laži imaju poten-

cijal širenja "dalje, brže, dublje i šire" od istine (Vosoughi, Roy i Aral 2018: 1147).

Širenje pogrešnih i namjerno manipuliranih informacija i dezinformacija *online*-platformama već je neko vrijeme prepoznato kao jedan od ključnih izazova suvremenim demokracijama (Europska komisija 2018). Pritisak međunarodnih organizacija i pojedinih nacionalnih vlasti, pogotovo unutar Europske unije, rezultirao je nizom mjera koje su vodeće tehnološke tvrtke primijenile u svojim uslugama kako bi smanjile problem. No pandemija koronavirusa povećala je zabrinutost te su Svjetska zdravstvena organizacija (2020) i Europska komisija (2020) dodatno pojačale suradnju s platformama u borbi protiv infodemije. Sve aktivnosti koje platforme poduzimaju dobrovoljne su, budući da djelovanje platformi na tom području gotovo uopće nije regulirano, pogotovo na nadnacionalnoj razini.

U ovome radu analiziraju se politike i tehnološka rješenja koja su vodeće *online*-platforme primijenile u jeku pandemije bolesti COVID-19 odgovarajući na problem širenja dezinformacija kao koordinirane aktivnosti, kao i nepovjerenih informacija iz različitih izvora. Cilj rada je rasvijetliti nove pristupe u prikazu, moderaciji i rangiranju sadržaja te se kritički osvrnuti na potencijalne nedostatke tih politika i implikacije koje mogu imati za informirano građanstvo i pluralističnu javnu raspravu. Analiza se temelji na teoriji participativne demokracije koja zahtijeva informirano građanstvo (Helberger 2019) koje može nastati samo ako postoji medijski pluralizam koji građanima jamči pristup različitim i vjerodostojnim izvorima informacija i mišljenja kako bi izgradili svoj stav, bez pretjeranog utjecaja dominantnog središta moći (EK 2007: 5; Valcke i dr. 2009: 2). U takvu sustavu jedno od temeljnih prava jest pravo na slobodu izražavanja te transparentnost

i opravdanost svake vrste intervencije u slobodu govora.

Prije izlaganja metodologije istraživanja i same analize, prikazat ću trenutačne rasprave o demokratskoj funkciji *online*-platformi i dati kraći pregled fokusa i učinkovitosti samoregulacije platformi u borbi protiv dezinformacija.

## **Demokratska uloga *online*-platformi**

Profesionalni i neovisni mediji odavno su prepoznati kao jedna od ključnih sastavnica demokracije. S jedne strane, građanima osiguravaju informacije koje su im nužne za participaciju i donošenje odgovornih i informiranih odluka. S druge strane, mediji u demokraciji djeluju kao četvrta vlast ili "psi čuvari" koji nadziru djelovanje poluga vlasti i moći u društvu. Mediji mogu ispunjavati te uloge samo ako je osiguran pluralizam, jedan od glavnih ciljeva medijskih politika koji se ostvaruje učinkovitom zaštitom slobode izražavanja i prava na pristup informacijama, zaštitom profesionalnih uvjeta i sigurnosti novinara u obavljanju posla, transparentnošću i pluralnošću medijskog vlasništva, političkom neovisnošću medija i društvenom uključivošću, odnosno primjerenim predstavljanjem različitih interesa i svjetonazora, kako u upravljačkim strukturama tako i u sadržajima (Brogi i dr. 2018).

Tijekom posljednja dva desetljeća broj izvora i opseg vijesti dostupnih *online* dramatično su se povećali. Tome su u velikoj mjeri pridonijele *online*-platforme koje su omogućile pojedincima da dijele informacije s potencijalno globalnim publikama. Iako bi se to na prvi pogled moglo tumačiti kao demokratizacija i veća pluralizacija u javnoj komunikaciji, dokazi sugeriraju kako je možda ipak riječ o previše informacija koje vode do preopterećenja vijestima (Holton i Chyi 2012; Newman i dr. 2019) i informacij-

skog poremećaja (Wardle i Derakhshan 2017), a što potencijalno ima negativne posljedice za informirano građanstvo i demokraciju.

*Online*-platforme, poput društvenih medija, ne proizvode vlastiti sadržaj, ali algoritamski kreiraju personalizirane preporuke medijskoga, političkog i drugog sadržaja prema obilježjima i preferencijama pojedinačnih korisnika. *Facebook*, primjerice, ne distribuira samo sadržaj nego kreira i arhitekturu u kojoj korisnici pronalaze i ostvaruju interakciju s vijestima (Helberger 2016). Helberger to objašnjava kao sposobnost toga društvenog medija da pozornost velikog broja korisnika usmjeri na određena važna pitanja, a da druge teme gurne u zaborav. Na taj način *Facebook* oblikuje uvjete javne rasprave, što ga čini "društvenim urednikom" i "privatno kontroliranom javnom sferom", koja pokazuje nov oblik vertikalne integracije u kojemu prikupljanje i korištenje velikog broja podataka o korisnicima ima središnju ulogu (Helberger 2016).

U jednom od ključnih radova o demokratskoj ulozi *online*-platformi i posebno njihovih sustava za preporuku vijesti, Helberger (2019) tvrdi da se demokratska uloga medija danas ne može ostvariti bez promišljanja demokratske uloge algoritamskih sustava za preporuku sadržaja. Ta se uloga može promatrati iz perspektive različitih teorija demokracije: od liberalne demokracije, koja u središte stavlja individualne slobode i osobnu autonomiju u izboru informacija prema vlastitim preferencijama, do participativne demokracije, u kojoj građani trebaju biti izloženi ne samo onome što "žele znati" nego i onome što "moraju znati" kako bi mogli sudjelovati u političkim procesima (Helberger 2019). Dok su u prvoj fazi postojanja *online*-platforme i njihovi algoritamski sustavi za preporuku sadržaja bili bliži liberalnom modelu, stavljajući u središte interese i informa-

cijske preferencije pojedinačnih korisnika na temelju podataka i saznanja koja o njima prikupljaju, u novije vrijeme sve je veći naglasak na očekivanju da platforme moderiraju i prezentiraju sadržaj tako da to odražava raznolikost ideja i mišljenja u društvu. Ukratko, od platformi se sve više očekuje da, kao jedne od ključnih aktera u informacijskom okruženju, ostvaruju svoju demokratsku funkciju iz perspektive participativne demokracije koja se temelji na načelima pluralizma. Helberger (2019) naglašava kako takav pristup uključuje transparentnosti i odgovornosti onih koji kreiraju algoritamsko donošenje odluka jer je pozornost ljudi smanjena, pa kriteriji prema kojima nešto dobiva veću, a nešto manju vidljivost moraju biti jasni i provjerljivi, a trenutačno nije tako.

U tradicionalnim medijima urednička je odgovornost zakonski regulirana. Kod platformi, koje nisu mediji u tradicionalnom smislu, ali nisu ni isključivo tehnološka poduzeća jer sve više djeluju kao *gatekeeperi* (Powers 2017) i kao arene za političku debatu (Woolley i Howard 2016), slična regulacija uredničke odgovornosti, pluralizma i zahtjeva za transparentnošću još uvijek uglavnom ne postoji. Njihove funkcije, koje su slične medijskima, trenutačno nisu pod nadzorom medijskih i drugih regulatora u većini zemalja u kojima djeluju. Jedno je od otvorenih pitanja jurisdikcija, budući da je uglavnom riječ o američkim poduzećima i korporacijama, koja znatno oblikuju informacijska okruženja diljem svijeta. Ne postoji konsenzus o tome tko bi, na koji način i na kojoj razni trebao regulirati djelovanje platformi s obzirom na to da ono zadire u različita područja javnih i komercijalnih aktivnosti (Helberger, Pierson i Poell 2018) te da se odnos društvenih i tržišnih vrijednosti drugačije interpretira u Sjedinjenim Američkim Državama nego, primjerice, u Europi (Gillespie 2018), a vrijednosne

razlike postaju još izraženijima kada se sagleda globalna slika. I same platforme koriste taj "poseban status" kako bi zabilazile sektorske regulative i javnu odgovornost (Van Dijk 2020).

Jasno je da regulacija i politike dizajnirane za tradicionalne medije ne odgovaraju kompleksnosti izazova povezanih s *online*-platformama. Razlika je posebno očita u dimenziji prikupljanja i korištenja velikog broja podataka o korisnicima, što platformama daje neviđenu prednost u količini znanja i kontroli koju imaju nad korisnicima, kao i u mogućnostima za oblikovanje informacijskog okruženja (Helberger, Kleinen-von Königslöw i Van der Noll 2015) te u privlačenju oglašivača kroz mikrotargetirano oglašavanje koje je u određenim slučajevima glavni katalizator širenja dezinformacija (Nenadić 2019). Usprkos tome što imaju vitalnu ulogu u ostvarivanju važnih javnih vrijednosti, djelovanje platformi trenutačno je gotovo potpuno prepušteno samoregulaciji utemeljenoj na njihovoj dobrovoljnoj suradnji s međunarodnim organizacijama, poput Europske komisije i Svjetske zdravstvene organizacije. Takva vrsta samoregulacije povećava ionako veliku moć platformi u odlučivanju o vjerodostojnosti, vrijednosti i legalnosti sadržaja, a istodobno nema učinkovite transparentnosti njihova djelovanja, kako u donošenju internih politika kojima oblikuju informacijsko okruženje tako i u njihovoj implementaciji.

### **Samoregulacija platformi u borbi protiv dezinformacija**

Djelovanje *online*-platformi temelji se uglavnom na njihovim internim politikama, tehnologijama i uvjetima korištenja. Gotovo da nema regulacije na tom području, a upitno je bi li regulacija sadržaja donijela više koristi ili štete, ako se u obzir uzme kompleksnost implementacije budući da je riječ o korpo-

racijama koje djeluju globalno i posreduju goleme količine sadržaja, a da se pritom ne naštetiti slobodi izražavanja. Posljednjih godina, pod pritiskom zainteresiranih javnosti i razvojem iskustva i metodologija znanstvene zajednice, politike moderacije sadržaja vodećih platformi postale su nešto jasnije, ali još nisu potpuno transparentne. Primjerice, još ne znamo točno koje sve kriterije *Facebookov* algoritam uzima u obzir u određivanju i rangiranju sadržaja prema svakome pojedinom korisniku, a nije ni potpuno jasno odvaja li *Facebook* i kako to čini političko oglašavanje od drugih oglasa koji nisu izravno politički, ali promiču određenu političku agendu.

Barrett i Kreiss (2019) označili su politike moderacije sadržaja vodećih platformi "nestalnim". Kada govore o "nestalnim platformama", oni ukazuju na to da se njihove politike mijenjaju u kratkim razdobljima, često i dramatično, i to uglavnom kao rezultat vanjskih pritisaka i ekonomskih razloga. To se događa u tolikoj mjeri da je promjene gotovo nemoguće pratiti, pogotovo kada se uzme u obzir da vodeće platforme često ne primjenjuju sve novine na svim tržištima na kojima djeluju ili to makar ne čine istodobno. Osim toga, određene politike, iako su implementirane na jednak način, mogu imati potpuno drugačije implikacije u različitim kulturnim i političkim okruženjima.

Europska komisija objavila je 2018. "Komunikaciju" u kojoj problem širenja velikog broja dezinformacija *online* navodi kao jedan od glavnih izazova demokraciji. Tim dokumentom najavila je i svoj intervencionistički pristup platformama (Flew i dr. 2019; Nenadić 2019) te postavila temelje Kodeksa Europske unije o suzbijanju dezinformacija. Kodeks je predstavljen kao samoregulatorni instrument koji bi trebao osigurati proaktivnost platformi u ograničavanju širenja dezinformacija, odnosno trebao bi

biti prijelazna mjera kojom se privatnim akterima povjerava važan zadatak povećanja transparentnosti i vjerodostojnosti informacijskog okruženja. Ako platforme u tome ne uspiju, samoregulaciju bi mogla zamijeniti regulacija (EK 2018b). Kodeks su dosad potpisali *Facebook*, *Google*, *Twitter*, *Mozilla*, *Microsoft* i *Tik Tok* te oglašivači i oglašivačka industrija. Obvezali su se da će: (1) nadzirati i prekinuti monetizaciju računa koji se lažno predstavljaju i šire dezinformacije; (2) povećati transparentnost političkog oglašavanja označavanjem sponzora i utrošenih sredstava; (3) označavati automatizirane račune (botove); (4) osigurati veću vidljivost vjerodostojnim sadržajima; (5) omogućiti istraživačkoj zajednici pristup podacima kako bi mogla analizirati problem širenja dezinformacija *online*.

Nakon prvih dvanaest mjeseci implementacije Kodeksa, Europska komisija (2020) objavila je sveobuhvatnu analizu njegove učinkovitosti. U analizi je Kodeks načelno prepoznat kao vrijedan instrument u borbi protiv dezinformacija, ali sa značajnim nedostacima koje bi trebalo otkloniti, uključujući neusklađenost definicija među platformama (primjerice, što je političko oglašavanje), različite strukture izvještaja, što otežava komparativnu evaluaciju, te još uvijek prisutan nedostatak transparentnosti prema istraživačkoj zajednici i regulatorima, budući da nije osiguran funkcionalan pristup podacima koji bi omogućio neovisan nadzor nad primjenom različitih mjera.<sup>1</sup> Osim toga, kao

<sup>1</sup> Na problem je ukazala i ERGA, Skupina europskih regulatora za audiovizualne medijske usluge, u svojem izvješću zasnovanom na analizi implementacije Kodeksa Europske komisije o suzbijanju dezinformacija. European Regulators Group for Audio-visual Media Services (2020). Report on Disinformation: Assessment of the implementation of the Code of Practice. <https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>

značajan nedostatak navodi se činjenica da je Kodeksu pristupio samo manji broj platformi i da ne uključuje, primjerice, aplikacije za razmjenu poruka poput *Messengera* ili *WhatsAppa*, iako one imaju sve veću ulogu u širenju dezinformacija (Elias i Catalan-Matamoros 2020). Stoga su sve glasnije sugestije da samoregulaciju zamijeni koregulacija ili tradicionalna regulacija (ERGA 2019). I Europska komisija razmatra kako određene dijelove Kodeksa ugraditi u paket Zakona o digitalnim uslugama (Digital Service Act, DSA), koji se trenutačno izrađuje kako bi zamijenio dvadeset godina staru Direktivu o e-trgovini.

No, dok se taj zakon počne primjenjivati, proći će godine. U međuvremenu, uključujući i borbu protiv pogrešnih i dezinformacija o koronavirusu, Europska komisija računa na pojačan, koordiniran i transparentniji angažman platformi kroz samoregulaciju. Zajednička komunikacija Europske komisije (2020), naslovljena "Borba protiv dezinformacija o bolesti COVID-19 – prepoznavanje činjenica", preuzela je termin Svjetske zdravstvene organizacije ističući kako je "infodemija" vezana za bolest COVID-19 globalan izazov i kako je suradnja s platformama bitan element učinkovitog odgovora na to. Europska komisija zatražila je od platformi da "intenziviraju svoje aktivnosti za suzbijanje rizika prouzročenih krizom" te da o provedenim mjerama redovito izvještavaju. Jedna od ključnih očekivanih mjera jest promicanje i osiguravanje veće vidljivosti informacija iz nacionalnih i međunarodnih zdravstvenih organizacija te profesionalnih medija.

Informacije koje kruže uključuju i obmanjujuće zdravstvene informacije s lažnim tvrdnjama ("pranje ruku ne pomaže", "koronavirus je opasan samo za starije osobe" i sl.). Takav sadržaj nije nužno nezakonit, ali može izravno ugroziti živote i ozbiljno naštetiti napo-

rima da se pandemija ograniči. Jedan od mogućih odgovora jest da ključni akteri, kao što su javna tijela, mediji, novinari, provjeravatelji činjenica i udruge civilnog društva, stave na raspolaganje lako dostupne i vjerodostojne sadržaje te da platforme takve sadržaje učine vidljivijima kroz svoje usluge.

U veljači 2020. predstavnici Svjetske zdravstvene organizacije sastali su se s predstavnicima trideset poduzeća u Sicilijskoj dolini kako bi dogovorili strategiju borbe protiv "infodemije",<sup>2</sup> što je rezultiralo značajnim promjenama u moderiranju sadržaja vodećih platformi. Prvi put platforme, kao privatne kompanije, ocjenjuju sadržaje i izvore na temelju "vjerodostojnosti", što oblikuje informacijsko okruženje u pandemiji i može imati dalekosežne posljedice nakon što pandemija završi. U nastavku analiziram ključne politike koje su platforme primijenile kako bi suzbile problem dezinformacija o koronavirusu i bolesti koju uzrokuje. No, prije toga, objasniti ću metodu i analitički uzorak.

## Metodologija

*Facebook* i *Alphabet* (*Google*), prema različitim izvorima, najutjecajnije su tehnološke korporacije s najpopularnijim uslugama. *Facebook* to ostvaruje kroz društvene mreže *Facebook* i *Instagram*, a *Alphabet* kroz tražilicu *Google* i platformu za dijeljenje videa *YouTube*.<sup>3</sup> *Twitter* je posebno popularan među tvorcima javnog mnijenja, poput novinara i političara (Newman 2020). *Facebook*,

<sup>2</sup> Više pojedinosti o aktivnosti Svjetske zdravstvene organizacije na: <https://www.who.int/news-room/detail/29-06-2020-covid-timeline>

<sup>3</sup> V., primjerice, Statista (2020) Most popular social networks worldwide as of July 2020, ranked by number of active users. Global social networks ranked by number of users 2020. na: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Google i Twitter bili su među prvim potpisnicima Kodeksa Europske unije o suzbijanju dezinformacija. Tih pet platformi – Facebook, Instagram, Google Search, YouTube i Twitter – stoga čini uzorak ovoga istraživanja. Metoda istraživanja je analiza dokumenata koje su odabrane platforme objavile od početka korona krize do 31. srpnja 2020. kako bi komunicirale promjene svojih politika moderacije sadržaja kao posljedice koronakrize, a uključeni su i izvještaji, koje su u kolovozu 2020. dostavile Europskoj komisiji, o mjerama koje su primijenile da bi ograničile širenje dezinformacija o bolesti COVID-19. Ukupno, riječ je o 16 objava i izvještaja koji prikazuju aktivnosti i alate koje su platforme primijenile kako bi ograničile širenje lažnih i obmanjujućih informacija kroz svoje usluge (Prilog 1). Glavne kategorije analize izražavaju ključne domene politika platformi u borbi protiv "infodemije", a proizlaze iz dogovorenih dimenzija Kodeksa Europske komisije o suzbijanju dezinformacija: (1) prominentnost autoritativnih izvora, (2) degradiranje ili uklanjanje štetnih i opasnih sadržaja, (3) sprječavanje koordiniranoga neautentičnog ponašanja i (4) politike oglašavanja.

## Analiza

Sve analizirane platforme primijenile su niz politika i tehničkih rješenja kako bi kroz sve četiri kategorije borbe protiv dezinformacija, usuglašene na temelju Kodeksa, odgovorile na izazove vezane za informacijski dio krize oko bolesti COVID-19. Rasprava koja slijedi strukturira i opisuje te intervencije, ali i kritički preispituje njihovu implementaciju i potencijalne nedostatke iz perspektive zaštite pluralizma i slobode izražavanja. Pritom treba imati na umu da među odabranim platformama postoje značajne strukturne razlike, kao i razlike u načinima na koje se koriste. Jednu platformu

različite skupine i pojedinci mogu koristiti s drugačijom svrhom (primjerice, Facebook može primarno služiti za zabavu i povezivanje s prijateljima, ali i kao primaran izvor informacija i vijesti ili kao poligon za politički marketing). Cilj prikaza poduzetih inicijativa nije kvantitativna valorizacija i usporedba djelovanja različitih platformi. Ponekad jedna inicijativa neke platforme može imati bolji učinak nego veći broj inicijativa druge platforme. Ovaj rad donosi pregled politika odabranih platformi u borbi protiv dezinformacija i infodemije kako bi se promislile posljedice koje te politike, izravno ili neizravno, kratkoročno ili dugoročno, mogu imati za medijski pluralizam i informirano građanstvo, uzimajući u obzir njihovo značenje u oblikovanju informacijskog okruženja.

### *Prominentnost autoritativnih izvora*

Povećanje vidljivosti autoritativnih i vjerodostojnih izvora jedna je od ključnih obveza na koju su pristali potpisnici Kodeksa Europske komisije o suzbijanju dezinformacija. Time su se obvezali da će "ulagati u tehnološka sredstva pomoću kojih se u pretragama, obavijestima o novom sadržaju ili drugim automatski rangiranim distribucijskim kanalima prednost, prema potrebi, daje relevantnim, izvornim i pouzdanim informacijama" (Kodeks II. D. 8.). Jedna od prvih inicijativa koju su sve analizirane platforme poduzele kako bi odgovorile izazovu infodemije bilo je upravo povećanje vidljivosti informacija koje potječu iz autoritativnih izvora, ponajprije od Svjetske zdravstvene organizacije i nacionalnih zdravstvenih institucija, ali i od vlada te javnih i drugih *mainstream*-medija.

Kao što je vidljivo iz tablice 1, većina tih inicijativa može se grupirati u tri kategorije: *pop-up*-obavijesti koje kori-

slike navode na informiranje o bolesti COVID-19 kod Svjetske zdravstvene organizacije ili službenih nacionalnih izvora; algoritamsko prioritiziranje i potenciranje autoritativnih izvora u pretraživanju i konzumiranju sadržaja o koronavirusu; algoritamsko prioritiziranje sadržaja koji su prošli proces provjere činjenica (*fact-checking*) ili su originalni (nisu reprodukcije). Platforme prilagođavaju inicijative prirodi svojih usluga te ih tijekom vremena, kao i prema dobitnim učincima dorađuju ili mijenjaju prema vlastitom nahođenju. Osim toga, nisu sve inicijative određene platforme dostupne u svim zemljama, kao što je također vidljivo iz tablice 1.

Iz perspektive zaštite pluralizma, kako tijekom krize COVID-19 tako i poslije nje, problem je to što nijedna od platformi ne definira precizno ni autoritativne i vjerodostojne izvore niti kriterije koje koristi u određivanju takvih izvora. Vrlo je važno u globalnoj zdravstvenoj krizi spriječiti širenje dezinformacija i raskrinkavati zablude promoviranjem vjerodostojnih izvora. No bitno je – ako takve odluke i procese oblikuju privatni akteri koje ne podliježu javnom nadzoru i nemaju uredničku ili drugu odgovornost – osigurati funkcionalnu transparentnost intervencija.

Institucionalni izvori nisu nepogrešivi, kao što ni vlade i središta političke moći nisu uvijek vjerodostojni. Primjeri su komuniciranja američkog predsjednika Donalda Trumpa i brazilskog predsjednika Jaira Bolsonaro. Bolsonaro su *Twitter*, *Facebook*, *Instagram* i *YouTube* uklonili objave zbog širenja problematičnih sugestija, poput odgovaranja ljudi od držanja fizičkog razmaka ili preporuka da se koristi hidroksiklorokin u liječenju COVID-19 (Wagner 2020). Trumpu je *Facebook* uklonio, a *Twitter* skrio iza upozorenja o "širenju obmanjujućih i potencijalno štetnih informacija" obja-

vu u kojoj tvrdi da je COVID-19 mnogo manje smrtonosan od gripe (BBC 2020).

Ako platforme nisu transparentne glede kriterija i procedura koje primjenjuju u određivanju vjerodostojnosti izvora, postoji rizik da se smanji vidljivost drugačijih glasova koji legitimno i vjerodostojno komuniciraju o istoj temi i nadziru rad institucija moći u svrhu zaštite javnog interesa. Posebno je važno da su platforme transparentne glede kriterija koje primjenjuju u određivanju i rangiranju vjerodostojnih medijskih izvora, budući da to ima izravne posljedice za medijski pluralizam.

Sustavi personaliziranih preporučitelja primjenjuju se na različitim područjima (primjerice, u *online*-kupovini), ali način na koji funkcioniraju u domeni distribucije vijesti i informativnih sadržaja sve više utječe na informacijske odabire pojedinaca (Helberger i dr. 2018), a time i na demokratsko građanstvo (Özgöbek i dr. 2014). Metodologija i logika algoritamskih preporučitelja sadržaja ni dosad nisu bile transparentne, osim nekih općih kriterija, poput recentnosti vijesti i izraženih preferencija korisnika (Özgöbek i dr. 2014). Iako su se još 2018, kroz Kodeks Europske unije o sprječavanju širenja dezinformacija, platforme obvezale na promoviranje "vjerodostojnih informacija", inzistirale su na tome da one ne bi smjele biti "suci istine" (Rosen 2019). *Facebook* je, primjerice, odgovornost za odlučivanje o vjerodostojnosti povjerio svojim korisnicima, pa je od 2018. *news feed*-algoritam u odabranim zemljama prioritizirao "vijesti iz publikacija koje korisnici ocjenjuje pouzdanima" (Mosseri 2019). To se početkom 2020. promijenilo. Pod utjecajem infodemije i pritiskom Svjetske zdravstvene organizacije, Europske komisije i drugih međunarodnih aktera, platforme su napravile dramatičan pomak – od odbijanja da budu arbitri istine do toga da su postale upravo to,



**Tablica 1.** Inicijative za povećanje vidljivosti autoritativnih izvora

	Google Search	YouTube	Facebook	Instagram	Twitter
<b>Inicijativa 1.</b>	Pretraživanje pojmova vezanih za koronavirus rezultira specijaliziranim pregledom <b>informacija i vizualizacija</b> na temelju službenih izvora (SZO)	Pretraživanje pojmova vezanih za koronavirus rezultira <i>pop-up</i> -poveznicom na <b>sadržaj SZO-a</b>	Pretraživanje pojmova vezanih za koronavirus rezultira <i>pop-up</i> -poveznicom na <b>zdravstvene institucije</b>	Pretraživanje pojmova vezanih za koronavirus rezultira <i>pop-up</i> -poveznicom na <b>zdravstvene institucije</b>	COVID-19 podsjetnik u pretraživanjima ( <i>pop-up</i> )
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Nije jasno naznačeno	Nije jasno naznačeno	U 79 zemalja i na 29 jezika (u 17 zemalja Europske unije, ali ne i u Hrvatskoj)
<b>Inicijativa 2.</b>	Veća vidljivost <b>autoritativnih izvora</b> u pretraživanjima vezanima za " <b>osjetljive događaje i kontekst</b> "	Sva videa o koronavirusu imaju poveznicu na <b>sadržaj SZO-a</b>	Covid-19 informacijski centar		COVID-19 stranica događaja s najnovijim objavama iz "vjerodostojnih i autoritativnih" izvora – vlada, medija, civilnog društva – na lokalnim jezicima
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Sve zemlje Europske unije		Više od 30 zemalja, ali nisu jasno naznačene
<b>Inicijativa 3.</b>	Algoritam daje veću vidljivost sadržajima koji su prošli <b>fact-checking</b> -proces (takvi sadržaji imaju posebnu oznaku)	Algoritam daje veću vidljivost autoritativnim izvorima poput " <b>medija i zdravstvenih vlasti</b> " kada je pretraživanje vezano za zdravstveni pojam	Prikaz informacija SZO-a na vrhu <i>feeda</i> i članovima skupina koje se bave temom COVID-19 ( <i>pop-up</i> )	Prikaz informacija SZO-a na vrhu <i>feeda</i>	Verificiranje eksperata uz pomoć zdravstvenih institucija te prioritiziranje računa čija je e-mail adresa povezana s autoritativnom organizacijom ili institucijom
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Nije jasno naznačeno	Nije jasno naznačeno	Nije jasno naznačeno
<b>Inicijativa 4.</b>			Prioritiziranje originalnog izvještavanja (transparentno autorstvo)		
<b>Područje primjene</b>	Sve zemlje Europske unije		Sadržaji na engleskom jeziku		

instruirajući svoje algoritme da daju posebnu vidljivost i poboljšaju dostupnost određenih izvora koji su označeni kao mjerodavni i vjerodostojni, a da pritom nisu ponudile jasne definicije i kriterije prema kojima određuju vjerodostojnost. To je bilo prvi put da platforme ocjenjuju sadržaj i izvore na temelju vjerodostojnosti i tako oblikuju informacijsko okruženje u pandemiji. Ako se takve politike zadrže i nastave razvijati, mogu imati dalekosežne posljedice nakon što pandemija završi i kada interes javnosti ne bude više usredotočen na jednu temu, pa će biti i mnogo kompleksnije odrediti mjerodavne izvore.

### *Degradiranje ili uklanjanje štetnih sadržaja*

Uklanjanje i blokiranje objava predsjednika dviju država ogledni su primjeri mjera koje su platforme primijenile kako bi, uz promicanje "vjerodostojnih izvora", zaustavile širenje lažnih, štetnih i obmanjujućih objava. Tablica 2. sadržava pregled tih inicijativa. Posebno je vidljivo da su sve platforme dopunile svoje politike i uvjete korištenja zabranom objavljivanja sadržaja koji bi mogao proizročiti fizičku štetu ili narušiti politike javnog zdravlja. Kodeks Europske unije o suzbijanju dezinformacija definira na početku ključan koncept – dezinformaciju – sukladno Komunikaciji Europske komisije (2018) kao "dokazivo lažnu ili obmanjujuću informaciju koja je kumulativno (a) smišljena, iznesena i širi se radi stjecanja ekonomske koristi ili namjernog zavaravanja javnosti i (b) može naškoditi javnom interesu, jer uključuje ugrožavanje demokratskih političkih procesa i procesa donošenja politika, ali i javnog dobra, kao što je zaštita zdravlja građana Europske unije, okoliša i sigurnosti".

Ilegalni sadržaji, poput govora mržnje, predmet su zakonske regulacije država-članica Europske unije (Jurova

2016). Dezinformacije su kompleksniji problem, budući da je riječ o sadržaju koji je štetan, ali uglavnom nije ilegalan. Međutim, u pandemiji netočne informacije o zdravstvenoj zaštiti, prevenciji i liječenju bolesti mogu odvratiti ljude od donošenja odluka koje štite njihovo zdravlje, izlažući i njih i druge ozbiljnu riziku. Zbog toga su se *online*-platforme našle pod povećanim pritiskom da urade nešto kako bi zaustavile ili ograničile širenje netočnih i obmanjujućih sadržaja. Kako je prikazano u tablici 2, u središtu je glavnih inicijativa otkrivanje i uklanjanje informacija vezanih za COVID-19 koje proturječe uputama zdravstvenih vlasti ili znanstvenom konsenzusu te stoga mogu proizročiti štetu. Neke su platforme, poput *Twittera*, proširile definicije štete kako bi obuhvatile objave koje poriču znanstvene činjenice i preporuke globalnih ili lokalnih zdravstvenih vlasti, koje dijele savjete za liječenje i zaštitu koji su štetni za zdravlje, kao i savjete koji nisu odmah štetni, ali je poznato da su neučinkoviti, te manipulativne tvrdnje koje potiču na paniku, kaos i govor mržnje.

Platforme te inicijative, kao i moderaciju sadržaja općenito, temelje na svojim uvjetima korištenja ili standardima zajednice, kako ih često nazivaju. Budući da je riječ o uslugama koje su dostupne globalno, njihovi uvjeti i standardi ne odražavaju nužno određene nacionalne zakonodavne okvire. Primjerice, iz perspektive zaštite ljudskih prava i slobode izražavanja bilo bi vrlo zabrinjavajuće da platforme prihvate i globalno primijene standard sadržan u singapurskom zakonu iz 2019. koji omogućuje ministru da izdaju naredbe platformama o uklanjanju sadržaja koje smatraju lažnima, a pojedinačnim prijestupnicima prijete kaznom zatvora do deset godina (Nenadić 2019). Iako je problem širenja dezinformacija globalno prepoznat i postoji stanovita suglasnost o potrebi da se

određenima regulatornim intervencijama odgovori na njega, vodeće svjetske organizacije i stručnjaci koji djeluju na području zaštite slobode izražavanja upozoravaju na to da intervencije koje ciljaju primarno na sadržaj i platformama povjeravaju još više ovlasti u odlukama o uklanjanju sadržaja u vrlo kratkima vremenskim okvirima i bez javnog nadzora predstavljaju rizik za slobodu izražavanja (Kaye 2018, 2019).

Cilj je sadašnjih pristupa rješavanju problema dezinformacija identificirati i ukloniti sve lažne i obmanjujuće informacije, što je nemoguće zbog niza prepreka. Zbog golemih količina sadržaja, njihovo otkrivanje i uklanjanje uglavnom se temelje na sustavima strojnog učenja koji nisu nepogrešivi. S jedne strane, ti sustavi ne uspijevaju otkriti i ukloniti sve sporne sadržaje; s druge strane, budući da strojevi još uvijek slabije razumiju kontekst, ponekad uklanjaju i legitimne oblike izražavanja. Uz samostalnu, strojno vođenu detekciju, platforme se oslanjaju i na prijave korisnika, a neke surađuju i s organizacijama za provjeru činjenica. Međutim, recentne analize pokazuju da značajan broj prijava problematičnog sadržaja ostaje neobrađenim (CCDH 2020), a ako proces uklanjanja predugo traje, upitna je njegova svrha. Skupina istraživača sa Sveučilišta u Oxfordu pokazala je da je u razdoblju od listopada 2019. do lipnja 2020. *YouTubeu* trebao prosječno 41 dan da ukloni videozapise koji sadržavaju lažne podatke te da manipulativni videozapisi povezani s koronavirusom ne nalaze publiku putem samoga *YouTubea* nego uglavnom dijeljenjem na *Facebooku*, a *Facebook* je oznake upozorenja o lažnim informacijama postavio na manje od jedan posto videozapisa s pogrešnim informacijama koje su se na platformi dijelile (Knuutila i dr. 2020).

Neke su platforme zbog mjera protiv širenja koronavirusa<sup>4</sup> bile prisiljene raditi s manje ljudskih moderatora koji su, u određenoj mjeri, trebali nadzirati djelovanje automatiziranih procesa. Kako je *Facebook* ustvrdio u Izvješčaju o provedbi standarda zajednice,<sup>5</sup> objavljenom u kolovozu 2020, odsutnost ljudskih moderatora smanjila je količinu sadržaja koji su uspijevali obraditi te je, iako to nije eksplicitno navedeno u Izvješčaju, povećan rizik od pogrešnog djelovanja strojeva koji su uvelike ostali bez ljudskog nadzora.

Vjerojatno je najvažniji problem u toj vrsti moderacije sadržaja velikih razmjera to što odluke o štetnosti i legalnosti sadržaja, koje su dosad uglavnom donosili sudovi u precizno definiranu postupku, sada jednostrano donose privatne korporacije u vrlo kratkim rokovima i na temelju svojih promjenjivih pravila koja nisu nužno jednaka u svim zemljama u kojima pružaju svoje usluge. To je problematično iz perspektive zaštite slobode izražavanja, budući da djelovanje platformi ne podliježe gotovo nikakvom nadzoru javnih tijela ili regulatora, jer nema zakona koji ih obvezuje da svoje postupke učine transparentnima. Platforme dobrovoljno i regularno objavljuju svoje izvještaje, ali podaci su navedeni u agregiranu obliku, što onemogućuje detaljnije analize pojedinih slučajeva ili nacionalnih okolnosti, a nema ni mogućnosti neovisne provjere iznesenih činjenica.

### *Sprječavanje koordiniranoga neautentičnog ponašanja*

Kodeksom iz 2018. platforme su se obvezale da će u Europskoj uniji provo-

<sup>4</sup> V. <https://www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/>

<sup>5</sup> V. <https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020/>

**Tablica 2.** Inicijative za degradiranje ili uklanjanje štetnih sadržaja

	Google Search	YouTube	Facebook	Instagram	Twitter
<b>Inicijativa 1.</b>	Otkrivanje i uklanjanje informacija vezanih za COVID-19 koje protuslove uputama zdravstvenih vlasti i mogu prouzročiti štetu	Opći standardi: zabranjeni su sadržaji koji potiču opasne ili ilegalne aktivnosti i sadržavaju rizik od ozbiljne fizičke štete ili smrti, uključujući pogrešne medicinske informacije	<b>Politika protiv koordinirane štete:</b> uklanjanje sadržaja koji imaju potencijal prouzročiti fizičku/materijalnu štetu kroz (npr. poticanje na širenje bolesti, organiziranje okupljanja, "korona party" i sl.).  <b>Politika protiv lažnih informacija i štete:</b> uklanjanje sadržaja koji pridonose neizbježnom nasilju i fizičkoj šteti (lažne preventivne mjere ili neutemeljeno i opasno liječenje), suradnja sa SZO-om u identificiranju takvih lažnih tvrdnji.  <b>Politika protiv govora mržnje:</b> uklanjanje sadržaja koji tvrde da ljudi određenih obilježja (rasnih, vjerskih i dr.) imaju virus, da su stvorili virus ili da ga šire, kao i sadržaja koji se izruguje s određenim skupinama oboljelih od COVID-a-19		<b>Proširena definicija štete i zahtjev za uklanjanjem sadržaja koji: poriču znanstvene činjenice i preporuke globalnih ili lokalnih zdravstvenih vlasti; dijele savjete za liječenje i zaštitu koji su štetni za zdravlje, kao i onih koji nisu odmah štetni, ali je poznato da su neučinkoviti; manipulativne tvrdnje</b> koje potiču na paniku, kaos i govor mržnje
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Sve zemlje Europske unije, ali proaktivna tehnologija otkrivanja govora mržnje dostupna je samo na određenim jezicima (npr. engleskom i španjolskom)		Nije jasno naznačeno
<b>Inicijativa 2.</b>	Ispravak ili brisanje informacija koje se pojavljuju u pretraživanju pod oznakom "medicinskih tema", a suprotni su općemu znanstvenom konsenzusu	Posebna politika medicinskih informacija o COVID-u-19: uklanjaju se <b>pogrešne infomacije i dezinformacije</b> . Primjerice, zabranjeno je nijekanje postojanja bolesti ili poticanje na korištenje neodobrenih kućnih lijekova	<b>Politika reguliranih dobara:</b> uklanjanje sadržaja kojima se pokušava iskoristiti kriza za financijsku dobit, a može prouzročiti i štetu (npr. trgovanje maskama, dezinficijensima ili testovima na COVID-19)		Kako bi bio uklonjen, sadržaj mora zadovoljiti tri uvjeta: da je tvrdnja o činjenici (ne mišljenje); da je dokazivo lažan ili obmanjujući; da bi vjerovanje u tvrdnju prouzročilo štetu. Granični slučajevi mogu biti ostavljeni, uz upozorenje, a opetovano objavljivanje takva sadržaja može rezultirati blokadom računa
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Sve zemlje Europske unije		Nije jasno naznačeno
<b>Inicijativa 3.</b>		<b>Granični sadržaji</b> i sadržaji koji mogu na štetne načine dezinformirati korisnike bit će manje vidljivi u personaliziranu sustavu preporučivanja sadržaja	<b>Politika protiv maltretiranja i uznemiravanja:</b> uklanjanje sadržaja koji zlonamjerno cilja na pojedince tvrdeći, primjerice, da su oboljeli od COVID-19 (osim ako to nisu sami javno objavili)		<b>Označavanje i upozorenja:</b> dodatni kontekst i informacije, osim objava koje sadržavaju sintetički ili manipulirani multimedijski sadržaj; osim objava koje sadržavaju osporene ili obmanjujuće informacije povezane s COVID-19 + poveznica na vanjski pouzdan izvor o temi

**Tablica 2.** Inicijative za degradiranje ili uklanjanje štetnih sadržaja (*nast.*)

	Google Search	YouTube	Facebook	Instagram	Twitter
<b>Područje primjene</b>		Sve zemlje Europske unije	Nije jasno naznačeno, ali proaktivna tehnologija otkrivanja takva sadržaja dostupna je samo na engleskom jeziku		Nije jasno naznačeno
<b>Inicijativa 4.</b>			Kada je sadržaj uklonjen zbog povreda politika platforme, korisnik koji ga je objavio dobije obavijest, priliku za pritužbu i dodatne informacije kako bi "ispravio ponašanje u budućnosti". I onomu koji je objavio sadržaj i drugima koji su mu bili izloženi pokazuje se poruka koja ih vodi na stranicu SZO-a s razotkrivenim mitovima o COVID-19.		Proaktivan nadzor sadržaja vezanih za COVID-19.
<b>Područje primjene</b>			Sve zemlje Europske unije		Nije jasno naznačeno
<b>Inicijativa 5.</b>			<b>Smanjena vidljivost</b> sadržaja koji su <b>fact-checkeri</b> , koji surađuju s <i>Facebookom</i> , razotkrili kao lažan. Onima koji naidu na takav sadržaj daje se oznaka upozorenja.		
<b>Područje primjene</b>			Nije jasno naznačeno		

diti jasne politike označavanja i sprječavanja zlorabe automatiziranih računa i djelovanja kroz svoje usluge. Taj pristup uključuje primjenu sustava koji će omogućiti označavanja botova i jasno razlikovanje automatiziranog djelovanja od ljudske interakcije. Osim toga, platforme sve intenzivnije rade na otkrivanju slučajeva manipulacije njihovim uslugama te zloćudnih utjecaja ili koordiniranoga neautentičnog ponašanja. Inicijative na tom području borbe protiv infodemije (tablica 3) veći naglasak stavljaju na aktere koji zlorabe funkcionalnosti platformi i ponašanje koje se kosi s uvjetima korištenja i standardima zajednice od intervencija u sam sadržaj.

U izvještajima, koje su u kolovozu 2020. dostavile Europskoj komisiji, o mjerama koje su primijenile kako bi ograničile širenje dezinformacije pove-

zanih s bolešću COVID-19, platforme su istaknule kako, iako su otkrile velik broj lažnih informacija, nisu u svojim uslugama otkrile koordinirane operacije širenja dezinformacija. To ukazuje na to da koordinirane informacijske operacije zlorabe platformi kako bi se ostvarili određeni interesi nisu velik problem postojeće zdravstvene krize, ali ostaju velik problem u izborima i političkim borbama općenito. Istraživači sa Sveučilišta Oxford, Bradshaw i Howard (2019), u svojoj su studiji prezentirali dokaze o postojanju organiziranih kampanja za manipulaciju javnim mnijenjem putem *online*-platformi u 70 zemalja, što je porast od gotovo 50 posto prema 2018. godini. Pokazali su kako u svakoj zemlji postoji najmanje jedna politička stranka ili vladina agencija koja koristi društvene medije kako bi pokušala utjecati na javno mnijenje.

**Tablica 3.** Inicijative za otkrivanje koordiniranoga neautentičnog ponašanja

	Google Search YouTube	Facebook Instagram	Twitter
<b>Inicijativa 1.</b>	<b>Uklanjanje sadržaja i blokada računa</b> koji sudjeluju u koordiniranome neautentičnom ponašanju kako bi se ostvario neki utjecaj (izvještavanje o aktivnostima u tromjesečnim biltenima)	<b>Fokus na ponašanje (i aktere)</b> , a ne na sadržaj: koordinirani naponi manipulacije javnom raspravom kako bi se ostvario određeni cilj, često lažnim računima (mjesečno izvještavanje o aktivnostima)	<b>Automatizirano otkrivanje i blokiranje</b> (nakon dodatne provjere) <b>umjetnog pojačavanja dosega</b> dezinformacija na području javnog zdravlja (izvještavanje o aktivnostima)
<b>Područje primjene</b>	Sve zemlje Europske unije	Sve zemlje Europske unije	Nije jasno naznačeno

Recentni primjer iz SAD-a pokazuje kako koordinirana manipulativna kampanja može istodobno promovirati dvije agende: bacati sumnju na izborni proces uoči predsjedničkih izbora i umanjivati opasnost bolesti COVID-19 koristeći lažne i manipulativne objave. U tom slučaju, *Twitter* i *Facebook* poduzeli su mjere za obustavljanje i uklanjanje računa koji su sudjelovali u kampanji. No, kako naglašava Shannon McGregor sa Sveučilišta u Sjevernoj Karolini (2020), to nije bilo prvi put da koordinirane manipulacije, i to u izvedbi pametnih tinejdžera, uspješno zaobilaze mjere određeno vrijeme. McGregor upozorava i na to da platforme svaki put reagiraju drugačije, a ponekad uopće ne reagiraju. "To se događa zato što je crta između 'koordiniranog ponašanja' i regularnih aktivnosti kampanje, prema definicijama platformi, nejasna. Dvosmislenost i nedosljedna provedba, kao i slučajan način moderiranja političkog govora, predstavljaju prijetnje izbornom procesu" (McGregor 2020).

Ako platforme nemaju jasne definicije "lažnih računa" i "koordiniranoga neautentičnog ponašanja" te ako njihovo izvještavanje o primjeni tih politika nije potpuno transparentno, postoji rizik da će reagiranjem na problematično ponašanje i aktere, greškom ukloniti i autentične račune, što se događalo i što su i

same priznale,<sup>6</sup> a što ugrožava slobodu izražavanja. Posebno je problematično ako se takve greške dogode kod legitimnih političkih aktera tijekom izborne kampanje u kojoj je vrijeme jedan od najoskudnijih resursa i svako izbjivanje iz javne rasprave može imati značajne posljedice za politički proces i demokraciju.

### *Politike oglašavanja*

Kako bi zaustavile stranice koje dezinformiraju u ostvarivanju zarade korištenjem njihovih sadržaja, *online*-platforme su definirale niz politika i standarda, a neke su dodatno unaprijeđene u kontekstu izazova infodemije. Još na temelju Kodeksa Europske unije o sprječavanju širenja dezinformacija, platforme su se obvezale da će primijeniti mjere koje će ograničiti protok oglašavanja na *web*-stranicama trećih strana koje šire dezinformacije, uključujući one o COVID-19. Kako bi spriječile dezinformiranje, ali i zarađivanje na iskorištavanju javne zdravstvene krize, privremeno su zabranile oglase koji prodaju medicinske maske, dezinficijense i testove na COVID-19.

Potpisavši Kodeks, platforme su pristale na označavanje svih oglasa u skladu s pravnim okvirom Europske unije i nacionalnim propisima kako bi se jasno

<sup>6</sup> V., primjerice, <https://about.fb.com/news/2019/05/fake-accounts/>

razlikovale od uredničkog sadržaja i kako bi bilo jasno tko ih je i koliko platio. U Kodeksu je značajna pozornost dana transparentnosti političkog oglašavanja, ali i "tematskom oglašavanju" koje se bavi važnima i aktualnim društvenim temama te, kao takvo, nije komercijalno oglašavanje, ali nije ni strogo političko prema uskoj definiciji Kodeksa ("oglas za izbor ili protiv izbora nekog kandidata ili nekog rezultata referenduma u nacionalnim i europskim izborima"). Zdravstvo je dodano društvenim pitanjima koja se pojavljuju u "tematskom oglašavanju" krajem 2019. i od tada, primjerice na *Facebooku*, svi oglašivači u Europskoj uniji koji žele zakupiti oglase o toj temi moraju proći proces autorizacije kojim se potvrđuje njihov identitet i kreira oznaka s informacijom o tome tko je oglas platio.

Reguliranjem političkoga i "tematskog oglašavanja" platforme nedvojbeno zadiru u domenu slobode izražavanja, zbog čega je iznimno važna transparentnost u definiranju i postupku koji primjenjuju, pogotovo zato što različite platforme imaju različite politike, a nerijetko ista platforma ima drugačije politike u različitim zemljama. Javno dostupni repozitoriji političkoga i tematskog oglašavanja, na koje su se platforme također obvezale Kodeksom, trebali su riješiti problem transparentnosti i omogućiti javni nadzor nad oglašivačima i sadržajima oglasa, ali i nad implementacijama politika samih platformi. No, zbog problematičnosti određivanja jasne i operativne definicije "tematskog oglašavanja", koja ne bi zadirala u slobodu izražavanja, arhive oglasa u sadašnjoj provedbi uglavnom daju informacije samo o političkom oglašavanju, i to tako da to onemogućava provjeru učinkovitosti politika samih platformi (Leersen i dr. 2019).

Recentna istraživanja pokazuju da platforme nisu posebno učinkovite ni

u primjeni dugih politika, poput razbijanja lanca monetizacije operaterima *web*-stranica koji šire dezinformacije. Taylor, Neudert, Hoffmann i Howard (2020), na uzorku od 830 izvora informacija o bolesti COVID-19, uspoređivali su djelovanje i doseg profesionalnih izvora u odnosu prema izvorima lažnih informacija i dezinformacija. Analizirali su optimizaciju za tražilice (SEO) svake od tih stranica kao sredstvo procjene njihove *online*-reputacije i njihova oslanjanja na oglašavanje u postizanju veće vidljivosti i dosega. Rezultati pokazuju da su alternativni izvori koji šire dezinformacije bolje optimizirani za distribuciju kroz tražilice i društvene medije od profesionalnih medijskih izvora te da se pritom u značajnoj mjeri oslanjaju na vodeće oglašivačke platforme u monetizaciji svojih *web*-stranica. Više od 60 posto stranica koje su širile dezinformacije koristilo je *Google*ve oglase.

To je samo jedno od više istraživanja koje pokazuje kako širenje dezinformacija može biti i ekonomski unosno, a da platforme svojim mjerama ne uspijevaju uvijek učinkovito prekinuti tokove zarade takvih stranica. Istodobno, zbog nedostatka transparentnosti, postoje sumnje da se automatiziranom i brzom primjenom vlastitih politika ponekad mogu oštetiti legitimni akteri ili ukloniti legitimni sadržaji, što je rizično za slobodu izražavanja i medijski pluralizam.

## Zaključak

Prikazujući politike koje su vodeće platforme – *Facebook*, *Instagram*, *Google Search*, *YouTube* i *Twitter* – primijenile u borbi protiv dezinformacija i infodemije povezane s pandemijom bolesti COVID-19, rad razmatra posljedica koje te politike, izravno ili neizravno, kratkoročno ili dugoročno, mogu imati za medijski pluralizam i informirano građanstvo. To čini iz perspektive utjecaja koji *online*-platforme kao privatna

**Tablica 4.** Politike oglašavanja u krizi COVID-19

	Google Search YouTube	Facebook Instagram	Twitter
<b>Inicijativa 1.</b>	<p><b>Blokiranje i uklanjanje oglasa</b> koji krše pravila platforme i <b>suspenzija (ili trajno uklanjanje) računa</b> s kojih su ti oglasi zakupljeni, kao i blokada oglasa na (trećim) web-stranicama koje su u Googleovoj mreži oglašavanja. <b>AdSense</b>, platforma za oglašavanje koja omogućuje operaterima web-stranica da unovče svoje stranice prikazivanjem kontekstualno relevantnih Googleovih oglasa, generirajući prihod na temelju klikova ili pojavljivanja, zabranjuje monetizaciju sadržaja koji "promoviraju ili zagovaraju štetne medicinske tvrdnje ili prakse" (<b>Pravila o opasnom ili pogrđnom sadržaju</b>).</p> <p><b>GoogleAds</b> ima posebna <b>Pravila za osjetljive događaje</b> (kao što je kriza COVID-19) – zabrana oglašavanja koje potencijalno profitira na pandemiji</p>	<p><b>Kontroverzni sadržaji i zabrana eksploatacijskih taktika u oglašavanju:</b> privremena zabrana oglasa koji prodaju medicinske maske, dezinficijense i pribor za testiranja na COVID-19.</p> <p><b>Lijekovi:</b> zabrana oglašavanja lijekova i cjepiva</p>	<p><b>Oglašavanje</b> koje sadržava eksplicitnu ili implicitnu poveznicu s COVID-19 <b>ograničeno</b> je. Zabranjen je senzacionalistički sadržaj u oglasima i onaj koji može izazvati paniku; povećanje cijena zbog potražnje; oglašavanje medicinskih maski i dezinficijensa; spominjanje testova, cjepiva i terapija, osim kada je riječ o informaciji koju promoviraju mediji. Dopuslene su objave vlada i SZO-a od javnog interesa</p>
<b>Područje primjene</b>	Sve zemlje Europske unije	Nije jasno naznačeno	Nije jasno naznačeno
<b>Inicijativa 2.</b>	<p><b>Sufinanciranje ili besplatno plasiranje oglasa</b> i obavijesti o COVID-19 od strane vladinih i međuvladinih i nevladinih organizacija, zdravstvenih institucija i drugi računi koji imaju povijest poštovanja politika platforme i žele s javnošću podijeliti relevantne poruke</p>	<p><b>Lažne i obmanjujuće tvrdnje:</b> posebnim pravilom zabranjuju se oglasi koji sadržavaju obmanjujuće, lažne ili neutemeljene zdravstvene tvrdnje (primjerice, promocija korištenja nemedicinskih supstancija i tretmana u liječenju ili prevenciji bolesti COVID-19). Posebnim pravilom zabranjuju se oglasi čiji su sadržaj <b>fact-checkeri</b> razotkrili kao pogrešne ili lažne</p>	<p>Besplatni oglasni prostor nacionalnima i međunarodnim zdravstvenim organizacijama kako bi komunicirali poruke od javnog interesa</p>
<b>Područje primjene</b>	Sve zemlje Europske unije	Nije jasno naznačeno	Danska, Finska, Francuska, Njemačka, Italija, Španjolska
<b>Inicijativa 3.</b>		<p><b>Oglasi o zdravstvenim problemima:</b> od 2019. svi oglašivači u EU koji žele zakupiti oglase o društvenim pitanjima (uključujući zdravstvo), izborima i politici moraju proći proces autorizacije kojim se potvrđuje njihov identitet i kreira oznaka s informacijom tko je platio oglas. Oglašivači moraju imati lokalne predstavnike u zemljama u kojima žele plasirati oglas</p>	
<b>Područje primjene</b>		Sve zemlje Europske unije	



globalna poduzeća imaju na oblikovanje informacijskog okruženja i informiranje građana, na čemu se zasniva njihova sve izraženija uloga u demokratskim društvima. Pritom ne podliježu regulaciji koja štiti medijski pluralizam i slobodu izražavanja, poput tradicionalnih medija, nego se vode vlastitima jednostranim i promjenjivim standardima i uvjetima korištenja.

Analiza je strukturirana prema glavnim dimenzijama Kodeksa Europske unije o suzbijanju dezinformacija, na koje su se platforme obvezale 2018. i prema kojima su kreirale i svoje odgovore na infodemiju. Četiri ključne kategorije intervencija uključuju: (1) prominentnost autoritativnih izvora, (2) degradiranje ili uklanjanje štetnih i opasnih sadržaja, (3) sprječavanje koordiniranoga neautentičnog ponašanja i (4) politike oglašavanja. Analiza pokazuje kako su sve platforme primijenile politike povećavanja vidljivosti izvora koje su ocijenile autoritativnima i vjerodostojnima, uključujući algoritamske preporuke i rangiranja sadržaja kroz rezultate pretraživanja. Uz to, proširile su definiciju štete, pogotovo u odnosu prema javnom zdravlju, kako bi dodatnim mjerama degradirale ili uklonile lažne, oportune i obanjujućih objave. Osim procjene i djelovanja prema sadržaju, platforme su komunicirale i politike jasnijeg označavanja botova kako bi se osiguralo razlikovanje automatiziranog djelovanja od ljudske interakcije te kako bi se brže otkrilo i spriječilo koordinirano neautentično ponašanje. Budući da u podlozi takvoga ponašanja i zlorabab funkcionalnosti platformi mogu biti i primarno ekonomski interesi, platforme su primijenile dodatne mjere kako bi ograničile protok oglašavanja na

*web*-stranicama trećih strana koje šire dezinformacije, ali i transparentnije označavanje različitih vrsta oglasa kao i njihovih sponzora.

Iako su platforme nedvojbeno poduzele niz potencijalno korisnih mjera kako bi odgovorile na izazove širenja lažnih, obmanjujućih i štetnih informacija u pandemiji, gotovo svaka od tih mjera sadržava i potencijalan rizik za slobodu izražavanja i medijski pluralizam. Ti rizici ne proizlaze toliko iz samih inicijativa koliko iz nedostatka regulatornog konteksta i neravnoteže između odgovornosti i moći koju platforme posjeduju. Naime, trenutačno ne postoji regulatorni okvir koji bi osigurao transparentnost i odgovornost djelovanja tih, ponajprije privatnih, poduzeća i korporacija, kako na nacionalnoj tako i na nadnacionalnoj razini. To im omogućava da politike donose, mijenjaju i provode jednostrano, bez javnog nadzora neovisnih regulatornih tijela koja bi štitila javni interes i pluralizam. Potreba za brzom i učinkovitom reakcijom na dezinformacije u pandemiji dovela je do toga da su utjecajne međunarodne organizacije, poput Svjetske zdravstvene organizacije, Europske komisije i drugih, opet platformama dale veliku moć, a da nisu osigurale mehanizme za ostvarivanje i veće odgovornosti. To je rezultiralo time da platforme primjenjuju politike kojima rangiraju izvore na temelju vjerodostojnosti, a da pritom nije jasno koje kriterije i definicije u tom procesu primjenjuju. I to je rizično za medijski pluralizam i slobodu izražavanja, pogotovo zato što trenutačno ne postoji neovisno tijelo koje bi imalo zakonom određene ovlasti da osigura transparentnost politika moderacije sadržaja, kao i analizu učinka te moderacije na pluralizam.

## Literatura

- Barrett, Brigit; Kreiss, Daniel. 2019. Platform transience: changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review*. ( 8) 4. DOI: 10.14763/2019.4.1446
- BBC. 2020. Trump Covid post deleted by Facebook and hidden by Twitter. *BBC.com* 6. listopada. <https://www.bbc.com/news/technology-54440662> (pristupljeno 7. listopada 2020).
- Bradshaw, Samantha, Howard, Philip N. 2019. *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Working Paper 3. Oxford, UK: Project on Computational Propaganda.
- Broggi, Elda, Nenadić, Iva, Parcu, Pier Luigi, De Azevedo, Viola, Cunha, Mario. 2018. *Monitoring Media Pluralism in Europe: Application of the Media Pluralism Monitor 2017 in the European Union, FYROM, Serbia and Turkey*. (Technical Report). Centre for Media Pluralism and Media Freedom, European University Institute. [https://cmpf.eui.eu/wp-content/uploads/2018/12/Media-Pluralism-Monitor\\_CMPF-report\\_MPM2017\\_A.pdf](https://cmpf.eui.eu/wp-content/uploads/2018/12/Media-Pluralism-Monitor_CMPF-report_MPM2017_A.pdf)
- CCDH (Center for Countering Digital Hate). 2020. *Failure to Act: How Tech Giants Continue to Defy Calls to Rein in Vaccine Misinformation*. [https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9\\_8d23c70f0a014b-3c9e2cfc334d4472dc.pdf](https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_8d23c70f0a014b-3c9e2cfc334d4472dc.pdf)
- Code of Practice on Disinformation. 2018. <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>
- ERGA (European Regulators Group for Audiovisual Media Services). 2020. *Report on Disinformation: Assessment of the implementation of the Code of Practice*. <https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>
- ERGA (European Regulators Group for Audiovisual Media Services). 2019. *Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation*. Slovakia: European Regulators Group for Audiovisual Media Services. [https://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06\\_Report-intermediate-monitoring-Code-of-Practice-on-disinformation.pdf?fbclid=IwAR1BZV2xYlJv-9nOzYAGhxA8AA5q70vYx0VUNn-h080WvDD2BffWFM3js4wg](https://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06_Report-intermediate-monitoring-Code-of-Practice-on-disinformation.pdf?fbclid=IwAR1BZV2xYlJv-9nOzYAGhxA8AA5q70vYx0VUNn-h080WvDD2BffWFM3js4wg)
- EK (Europska komisija). 2020. Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: *Tackling COVID-19 disinformation – Getting the facts right*. JOIN/2020/8 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020JC0008>
- EK (Europska komisija). 2018a. *Tackling online disinformation: a European Approach, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. COM/2018/236. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>
- EK (Europska komisija). 2018b. *Action Plan against Disinformation. European Commission contribution to the European Council*. [https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018\\_en.pdf](https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf)
- EK (Europska komisija). 2007. *Media pluralism in the Member States of the*

- European Union. Commission Staff Working Document No. SEC 32. [https://ec.europa.eu/information\\_society/media\\_taskforce/doc/pluralism/media\\_pluralism\\_swp\\_en.pdf](https://ec.europa.eu/information_society/media_taskforce/doc/pluralism/media_pluralism_swp_en.pdf)
- Elias, Carlos, Catalan-Matamoros, Daniel. 2020. Coronavirus in Spain: Fear of 'Official' Fake News Boosts WhatsApp and Alternative Sources. *Media and Communication*. (8) 2: 462–466. DOI: 10.17645/mac.v8i2.321
- Flew, Terry, Martin, Fiona, Suzor, Nicolas. 2019. Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media and Policy*. (10) 1: 33–50.
- Gauch, Susan, Speretta, Mirco, Chandramouli, Aravind, Micarelli, Alessandro. 2007. "User Profiles for Personalized Information Access." *The Adaptive Web: Methods and Strategies of Web Personalization*. Brusilovsky, Peter, Kobsa, Alfred, Nejdl, Wolfgang. (ur.). Berlin: Springer, str. 54–89.
- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.
- Helberger, Natali. 2019. On the Democratic Role of News Recommenders. *Digital Journalism*. (7) 8: 993–1012. DOI: 10.1080/21670811.2019.1623700
- Helberger, Natali, Karppinen, Kari, D'Acunto, Lucia. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*. (21) 2: 191–207.
- Helberger, Natali, Kleinen-Von Königslöw, Katharina, Van der Noll, Rob. 2015. Regulating the new information intermediaries as gatekeepers of information diversity. *Info* (17) 6: 50–71.
- Helberger, Natali, Pierson, Jo, Poell, Thomas. 2018. Governing online platforms: From contested to cooperative responsibility. *The Information Society*. (34) 1: 1–14. DOI: 10.1080/01972243.2017.1391913
- Helberger, Natali. 2016. Facebook is a new breed of editor: a social editor. *Media Policy Blog*, 15. rujna 2016. <https://blogs.lse.ac.uk/medialse/2016/09/15/facebook-is-a-new-breed-of-editor-a-social-editor/>
- Holton, Avery E., Chyi, Hsiang Iris. 2012. News and the Overloaded Consumer: Factors Influencing Information Overload Among News Consumers. *Cyberpsychology, behavior and social networking*. (15) 11: 619–24.
- Jurova, Vera. 2016. *Code of Conduct—Illegal online hate speech Questions and answers*. Factsheet. European Commission. [https://ec.europa.eu/info/sites/info/files/code\\_of\\_conduct\\_hate\\_speech\\_en.pdf](https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf)
- Kaye, David. 2019. *Report of the Special Rapporteur to the General Assembly on online hate speech (A/74/486)*. [https://www.ohchr.org/Documents/Issues/Opinion/A\\_74\\_486.pdf](https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf)
- Kaye, David. 2018. *Report of the Special Rapporteur to the Human Rights Council on online content regulation (A/HRC/38/35)*. <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>
- Knuutila, Aleks, Herasimenka, Aliaksandr, Au, Hubert, Bright, Jonathan, Howard, Philip N. 2020. *COVID-related misinformation on YouTube: The spread of misinformation videos on social media and the effectiveness of platform policies*. Data Memo. Oxford, UK: Project on Computational Propaganda.
- Leerssen, Paddy, Ausloos, Jef, Zarouali, Brahim, Helberger, Natali, De Vreesse, Claes H. 2019. *Platform Ad Archives: Promises and Pitfalls (2019)*.

- Internet Policy Review*. (8) 4. DOI: 10.14763/2019.4.1421
- McGregor, Shannon. 2020. What Even Is 'Coordinated Inauthentic Behavior' on Platforms? *Wired* 17. rujna 2020. <https://www.wired.com/story/what-even-is-coordinated-inauthentic-behavior-on-platforms/> (pristupljeno 7. listopada 2020).
- Mosseri, Adam. 2018. Helping Ensure News on Facebook is from Trusted Sources. *Facebook Newsroom* 19. siječnja 2018. <https://about.fb.com/news/2018/01/trusted-sources/> (pristupljeno 12. rujna 2020).
- Menadić, Iva. 2019. Unpacking the "European approach" to tackling challenges of disinformation and political manipulation. *Internet Policy Review*. (8) 4. DOI: 10.14763/2019.4.1436
- Newman, Nic. 2010. *#UKelection2010, mainstream media and the role of the internet: how social and digital media affected the business of politics and journalism* Working Paper. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Social%2520Media%2520and%2520the%2520Election.pdf>
- Newman, Nic, Fletcher, Richard, Kalogeropoulos, Antonis, Kleis Nielsen, Rasmus. 2019. *Reuters Institute Digital News Report 2019*. Oxford: Reuters Institute for the Study of Journalism. <http://www.digitalnewsreport.org/>
- Newman, Nic, Fletcher, Richard, Kalogeropoulos, Antonis, Levy, David A. L., Kleis Nielsen, Rasmus. 2018. *Digital News Report 2018*. Oxford: Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf>
- Oulasvirta, Antti, Blom, Jan. 2007. Motivations in Personalisation Behaviour. *Interacting with Computers*. (20) 1: 1–16.
- Özgöbek, Özlem, Gulla, Jon Atle, Erdur, R. Cenk. 2014. *A survey on challenges and methods in news recommendation*. Proceedings of the 10th International Conference on Web Information Systems and Technologies (WEBIST'14), str. 278–285.
- PAHO (Pan American Health Organization). 2020. *Understanding the Infodemic and Misinformation in the Fight against COVID-19*. Factsheet br. 5. [https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic\\_eng.pdf?sequence=14](https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf?sequence=14)
- Pariser, Eli. 2011. *The Filter Bubble: What the Internet is Hiding from You*. UK: Penguin.
- Powers, Elia. 2017. My News Feed is Filtered?. *Digital Journalism*. (5) 10: 1315–1335. DOI: 10.1080/21670811.2017.1286943
- Rosen, Guy. 2019. Remove, Reduce, Inform: New Steps to Manage Problematic Content. *Facebook Newsroom*. <https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/> (pristupljeno 12. rujna 2020).
- Sunstein, Cass. R. 2009. *Going to Extremes: How Like Minds Unite and Divide*. New York: Oxford University Press.
- Sunstein, Cass. R. 2002. *Republic.com*. Princeton: Princeton University Press.
- SZO (Svjetska zdravstvena organizacija). 2020. *Novel Coronavirus (2019-nCoV)*. Situation Report – 13 (2. veljače 2020). [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-n-cov-v3.pdf?sfvrsn=195f4010\\_6](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-n-cov-v3.pdf?sfvrsn=195f4010_6)
- Taylor, Emily, Neudert, Lisa-Maria, Hoffmann, Stacie, Howard, Philip N. 2020. *Follow the Money: How the Online Advertising Ecosystem Funds COVID-19 Junk News and Disinform-*

- mation. Working Paper 1. Oxford, UK: Project on Computational Propaganda. Comprop.oii.ox.ac.uk
- Valcke, Peggy i dr. 2009. *Independent Study on Indicators for Media Pluralism in the Member States – Towards a Risk-Based Approach* [Report]. ICRI, K.U. Leuven; CMCS, Central European University, MMTC, Jönköping Business School; Ernst & Young Consultancy Belgium. [https://ec.europa.eu/information\\_society/media\\_taskforce/doc/pluralism/pfr\\_report.pdf](https://ec.europa.eu/information_society/media_taskforce/doc/pluralism/pfr_report.pdf)
- Van Dijck, José. 2020. Governing digital societies: Private platforms, public values. *Computer Law & Security Review*. 36. DOI: 10.1016/j.clsr.2019.105377
- Van Dijck, José, Poell, Thomas, De Waal, Martijn. 2018. *The Platform Society. Public values in a connective world*. Oxford: Oxford University Press.
- Vosoughi, Soroush, Roy, Deb, Aral, Sinan. 2018. The spread of true and false news online. *Science*. (359) 6380: 1146–1151. DOI: 10.1126/science.aap9559
- Wagner, Kurt. 2020. Facebook, Twitter, YouTube Remove Posts From
- Bolsonaro. *Bloomberg* 31. ožujka. <https://www.bloomberg.com/news/articles/2020-03-31/facebook-twitter-pull-misleading-posts-from-brazil-s-bolsonaro> (pristupljeno 8. listopada 2020).
- Wardle, Claire, Derakhshan, Hossein. 2017. *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Report No. DGI(2017)09. Strasbourg: Council of Europe. <https://firstdraftnews.org/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-de%CC%81sinformation-1.pdf?x56713>
- Woolley, Samuel C., Howard, Philip N. 2016. Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents — Introduction. *International Journal of Communication*. 10: 4882–4890. <https://ijoc.org/index.php/ijoc/article/view/6298/1809>
- Zuiderveen Borgesius i dr. 2016. Should We Worry About Filter Bubbles? *Internet Policy Review*. (5) 1. <https://ssrn.com/abstract=2758126>

## Prilog: Popis analiziranih objava i izvještaja

1. Google COVID-19 report – August 2020. First baseline report – Fighting COVID-19 disinformation Monitoring Programme. European Commission <https://ec.europa.eu/digital-single-market/en/news/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>
2. Facebook COVID-19 report – August 2020. First baseline report – Fighting COVID-19 disinformation Monitoring Programme. European Commission. <https://ec.europa.eu/digital-single-market/en/news/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>
3. Twitter COVID-19 report – August 2020. First baseline report – Fighting COVID-19 disinformation Monitoring Programme. European Commission. <https://ec.europa.eu/digital-single-market/en/news/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme>

4. Keeping People Safe and Informed About the Coronavirus. Facebook, 5.10.2020. <https://about.fb.com/news/2020/08/coronavirus/>
5. An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19. Facebook, 16.4.2020. <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>
6. Combating COVID-19 Misinformation Across Our Apps. Facebook, 25.5.2020. <https://about.fb.com/news/2020/03/combating-covid-19-misinformation/>
7. Introducing New Authenticity Measures on Instagram. Instagram, 13.8.2020. <https://about.instagram.com/blog/announcements/introducing-new-authenticity-measures-on-instagram>
8. COVID-19: Community Guidelines Updates and Protections. Instagram, n.d. <https://help.instagram.com/477434105621119>
9. Coronavirus disease (COVID-19) Google Ads policy updates. Google, 15.6.2020. <https://support.google.com/google-ads/answer/9811449?hl=en>
10. Coronavirus disease 2019 (COVID-19) updates. YouTube, 3.7.2020. [https://support.google.com/youtube/answer/9777243?p=covid19\\_updates&visit\\_id=637350749445453338-1911282129&rd=1](https://support.google.com/youtube/answer/9777243?p=covid19_updates&visit_id=637350749445453338-1911282129&rd=1)
11. COVID-19 Medical Misinformation Policy. YouTube, n.d. <https://support.google.com/youtube/answer/9891785?hl=en>
12. An update on our continuity strategy during COVID-19. Twitter, 16.3.2020. [https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html)
13. Inappropriate content. Twitter, Business, n.d. <https://business.twitter.com/en/help/ads-policies/ads-content-policies/inappropriate-content.html>
14. Updating our approach to misleading information. Twitter, 11.5.2020. [https://blog.twitter.com/en\\_us/topics/product/2020/updating-our-approach-to-misleading-information.html](https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html)
15. Coronavirus: Staying safe and informed on Twitter. Twitter, 3.4.2020. [https://blog.twitter.com/en\\_us/topics/company/2020/covid-19.html](https://blog.twitter.com/en_us/topics/company/2020/covid-19.html)
16. Advertising during COVID-19. Twitter, n.d. [https://marketing.twitter.com/en\\_gb/insights/twitter-covid19-advertising-policy](https://marketing.twitter.com/en_gb/insights/twitter-covid19-advertising-policy)

## Power without responsibility: the content moderation policies of the online platforms in the response to infodemic

**Abstract** The paper provides an overview of the policies enrolled by the leading online platforms in the response to misinformation and infodemic around the COVID-19 pandemic. It further discusses the consequences these policies may have on media pluralism and informed citizenship both directly or indirectly, as well as in the short or long term. The method is an analysis of documents published by selected platforms to communicate changes in their content moderation policies in the period between the start of the novel coronavirus crisis in February and 31 July 2020, and it includes also the reports submitted to the European Commission in August 2020 to report on measures platforms have implemented to limit the spread of COVID-19-related misinformation. The analysis suggests that the platforms have undoubtedly taken a number of potentially useful measures to address the challenges of false, misleading and harmful information in a pandemic, but almost every one of these measures carries also a potential risk to freedom of expression and pluralism. These risks do not arise primarily from the initiatives themselves but rather from a lack of regulatory context and strong imbalance between accountability and the power that the platforms hold.

**Key words** disinformation, infodemic, COVID-19, online platforms, content moderation policies, media pluralism

