Footage ng Protesta na may Sumisigaw nang Pabor kay Duterte

2025-050-FB-UA

Buong Desisyon sa Kaso

1. Paglalarawan at Kaalaman Tungkol sa Kaso

1. Pagkatapos maaresto ng dating Pangulo ng Pilipinas na si Rodrigo Duterte noong Marso 2025 upang harapin ang mga kasong isinampa sa kaniya sa International Criminal Court (ICC) sa Netherlands, ni-reshare ng isang Facebook user ang video ng isa pang user. Nagpapakita ang orihinal na video ng mga taong nagtitipon-tipon sa isang kalye habang may bitbit na mga karatula at watawat ng Serbia. May text na naka-overlay sa video na nagsasabing, "Netherland," at pinatutugtog ang makabayang awiting Tagalog na "Bayan Ko," sa background. Maririnig din ang audio ng mga taong paulit-ulit na sumisigaw ng "Duterte!" Sikat ang awiting "Bayan Ko" sa panahon ng pagpoprotesta laban sa batas militar noong 1980s laban sa dating diktador at pangulong si Ferdinand Marcos Sr. Mayroong caption sa wikang Ingles ang orihinal na post: "Mga tagasuporta sa Netherlands." Ang ni-reshare na Facebook post, na paksa ng kasong ito, ay mayroong tatlong emoji na nagmamakaawang mukha bilang caption. Lumalabas na ang post ay isang protesta sa Serbia na walang kinalaman sa pag-aresto kay Duterte, na dinagdagan ng mga caption at audio upang magmukha itong isang demonstrasyong pabor kay Duterte na naganap sa Netherlands.

2. Na-share ang orihinal na Facebook post nang daan-daang beses at nakatanggap ito nang humigit-kumulang 100,000 view mula noong na-post ito. Noong na-post ito, na-flag ng isang Meta classifier ang content bilang posibleng maling impormasyon. Pagkatapos ay isinama ng kumpanya ang content sa online na pilang ibinibigay sa mga third-party na fact-checking partner nito. Nag-viral ang mga katulad na bersyon ng video na ito sa Facebook noong Marso 2025; marami ang na-fact check at na-rate na Hindi Totoo ng mga fact-checker ng Meta. Dahil sa dami ng nasabing post at sa bilis ng pagkalat ng mga ito sa platform, hindi ma-rate ng mga fact-checker ang lahat ng post na nasa pilang ito, na maaaring kinabibilangan ng mga post na

katulad ng content na na-fact check na nila. Ayon sa Meta, hindi ni-rate ng mga fact-checker ang orihinal na post, at ang mga rating bilang “hindi totoo” na ipinataw sa mga katulad na post ay hindi awtomatikong naipataw rito. Kahit hindi na-rate, pansamantalang ipinakita ng Meta ang content ng kaso nang mas mababa sa Facebook feed ng mga user na wala sa US, kasama na rito ang mga nasa Pilipinas, batay sa mga senyales na nanghuhulang naglalaman ito ng maling impormasyon.

3. Pagkaraan ng ilang araw, iniulat ng isang user ang ni-reshare na post, ngunit ayon sa Meta, hindi isinapriyoridad ang ulat upang maipasuri sa mga tao. Hindi inaksyunan ng Meta ang ulat at isinara na ito. Pagkatapos maabisuhang hindi tinanggal ng Meta ang content sa platform, inapela ng nag-ulat na user ang desisyon ng Meta. Pinanindigan ng isang tagasuring tao ang paunang desisyong panatilihin ang content, kaya inapela ng nag-ulat na user ang desisyon sa Oversight Board. Noong pinili ng Board ang kasong ito, idinagdag ulit ng Meta ang post sa isang pila para sa pag-fact check, ngunit hindi ito ni-rate.

4. Naitala ng Board ang sumusunod na konteksto sa pagsasagawa ng desisyon nito.

5. Sa panahon ng kaniyang katungkulan bilang pangulo ng Pilipinas mula 2016 hanggang 2022, nagsagawa si Rodrigo Duterte ng kampanya laban sa ilegal na droga, na ginamitan ng labis na puwersa at pagpatay ng mga vigilante o nang labag sa batas. Ang International Criminal Court (ICC) ay nagpasimula ng preliminaryong pagsusuri sa usapin, na sinundan ng isang buong imbestigasyon. Noong Marso 11, 2025, dinala si Duterte sa Netherlands upang harapin ang mga krimen laban sa sangkatauhan sa ICC.

6. Ang kasalukuyang pangulong Ferdinand Marcos, Jr., na anak ng dating diktador, ang nanalo sa pambansang halalan noong 2022, at si Sara Duterte naman, na anak ni Rodrigo Duterte, ang nanalo bilang pangalawang pangulo. Hindi nagtagal, pagkatapos manalo sa halalan, nawasak ang alyansa sa pulitika sa pagitan nila. Noong 2024 at 2025, lumala pa ang tunggalian sa pulitika. Sa unang bahagi ng taong ito, naghain ng mga reklamo ng impeachment o pagpapatalsik sa puwesto laban sa Pangalawang Pangulong Sara Duterte. Noong Pebrero, bumoto ang Kapulungan ng mga Kinatawan ng Pilipinas na i-impeach si Sara Duterte, na sa huli ay hindi rin natuloy. Ang pag-aresto kay Duterte noong Marso ay ipinagpalagay ng ilan bilang resulta ng

kanilang away sa pulitika. Ayon sa Pangalawang Pangulong Sara Duterte, maituturing na "pagdukot" ang naganap na pag-aresto, habang para naman sa mga aktibistang tagasulong ng mga karapatang pantao at sa mga pamilya ng mga napatay na biktima ng digmaan kontra droga o drug war, isa itong tagumpay para sa paghingi ng pananagutan.

7. Pagkatapos maaresto ni Duterte, sumiklab ang mga protesta bilang pagsuporta kay Duterte sa Pilipinas at sa Netherlands. Nag-rally rin ang mga aktibista at kamag-anak ng mga biktima ng drug war upang panagutin ang maysala. Sa panahong ito, nag-viral sa social media ang maraming video na may footage ng isang rally sa Serbia na inilahad bilang protestang pabor kay Duterte sa Netherlands, katulad ng content na inaanalisa.

8. Tatlong organisasyong fact-checker na nakikipagtulungan sa Meta sa Pilipinas ang nag-rate sa ilan sa mga video na ito bilang "Hindi Totoo" batay sa mga senyales na nagpakita ang footage ng protesta ng isang protesta sa Serbia, hindi sa Netherlands, at wala itong kinalaman sa pag-aresto sa dating Pangulong Duterte. Inilathala rin ng mga fact-checker na ito ang kanilang ratings na "Hindi Totoo" sa kanilang mga kaukulang website. Ayon sa AFP, ang mga video na ito ay "di-wastong gumagamit ng mga walang kinalamang visual upang palakihin ang bilang ng mga tao" (PC-31360). Ang isa pang katulad na video, na umani ng mahigit tatlong milyong view sa social media, ay nagpakita ng isang prusisyon ng mga Katoliko na ipinalabas bilang rally na pabor kay Duterte.

9. Noong nagsumite ito ng pampublikong komento para sa kasong ito, na-fact check ng AFP ang "mahigit 30 hindi totoo o mapanlinlang na claim" kaugnay ng pag-aresto kay Duterte (PC-31360). Naitala ng Rappler, na isa pang fact-checker para sa Meta sa Pilipinas, na "maraming bersyon ng partikular na minanipulang video na ito" ang nag-viral sa Facebook bilang bahagi ng "operasyong paulit-ulit na magpalabas ng impormasyon" at bahagi ito ng "peke at organisadong kampanya" na layong "magmanipula sa opinyon at asal ng publiko nang pabor sa kanya" (PC-31349). Ayon sa mga ekspertong kinonsulta ng Board, mula pa noong 2016, ang mga fact-checker at ang mga grupo ng lipunang sibil na ang sumasalo ng pasaning labanan ang pagpapakalat ng disimpormasyon.

10. Sa mas malawak na pag-unawa, ang paggamit ng social media sa Pilipinas ay isa sa pinakamatataas sa buong mundo, kung saan ang

Facebook ang pinakamadalas na ginagamit na platform. Malaki ang papel na ginagampanan ng social media sa paghubog ng pampublikong diskurso tungkol sa mga kasalukuyang kaganapan. Halimbawa, sa pambansang halalan noong 2016, ang mga tagabuo ng estratehiya sa advertising at ugnayang pampubliko ay nagsilbi bilang "mga arkitekto ng naka-network na disimpormasyon" upang mapangasiwaan ang mga kampanya sa pulitika, kung saan nag-hire sila ng mga online influencer at troll upang magpunla ng mga partikular na naratibo sa pulitika. Sa pambansang halalan noong 2022, kumalat sa social media ang historical revisionism o pagrerebisa sa kasaysayan tungkol sa legasiya ng pamilyang Marcos, na pinaniniwalaan ng ilan na nakapag-ambag sa pagkapanalo ni Marcos Jr. sa pagkapangulo. Ipinaliwanag ng mga ekspertong kinonsulta ng Board na ang "mga social media influencer, vlogger, at digital worker mula sa mga kumpanya ng advertising at ugnayang pampubliko ay may ginampanang malalaking papel sa paggawa at pagpapakalat ng disimpormasyon sa mga social media platform." Sa pagtindi ng tunggalian sa pagitan nina Pangulong Marcos at Pangalawang Pangulong Sara Duterte na nagdulot ng magkakasalungat na naratibo sa pulitika online, lalo nang nabahala ang mga Pilipino tungkol sa maling impormasyon, kung saan nagkaroon ng pag-aaral na nagsiwalat ng "rekord na 67%" ng mga tao ang tumindi ang pagkabahala tungkol sa maling impormasyon sa gitna ng taong 2025. Sa unang bahagi ng taong ito, nagpasimula ang Kongreso ng Pilipinas ng isang pagsisiyasat upang matugunan ang disimpormasyon sa social media, kung saan nag-imbita sila ng mga influencer, vlogger at kinatawan ng mga kumpanya ng social media upang magbigay ng salaysay.

11. Sa ilalim ng nakaraang pamahalaang Duterte at ng kasalukuyang pamahalaang Marcos, nagkaroon ng malulubhang hamon sa kalayaan sa pagpapahayag sa Pilipinas. Ang mga mamamahayag at media outlet na mapamuna kay Duterte ay inatake o ipinasara ng pamahalaan sa ilalim ng kaniyang pamumuno bilang pangulo at pinuntirya ng mga online troll. Ang "red-tagging," o ang pagtawag sa mga grupo o indibidwal bilang "mga tagasuporta, recruiter o miyembro ng New People's Army o ng Communist Party of the Philippines" ay naging pangkaraniwan nang gawain, na kadalasang namumuntirya sa mga mamamahayag at mga tagapagtanggol ng mga karapatang pantao. Nagpatuloy ang mga banta laban sa media sa ilalim ni Pangulong Marcos Jr. Noong Hunyo 2025, kinalampag ng UN Special Rapporteur sa kalayaan sa pagpapahayag ang pamahalaan ng Pilipinas upang tugunan ang pananakot at panggigipit ng mga mamamahayag at mga

tagapagtanggol ng mga karapatang pantao sa bansa. Ang mga hamong ito, kasabay ng pagdami ng mga kampanya sa disimpormasyon sa social media, ay nakasira sa tiwala ng publiko sa mga institusyon ng media sa Pilipinas.

2. Pagsusumite ng mga User

12. Ayon sa nag-uulat na user, sinadya ng content ng kaso na magpakita ng rally na sumusuporta kay Duterte kasunod ng pagkakakulong niya sa ICC, ngunit sa katunayan, hindi naman ito naganap. Ayon sa user, ibang kaganapan ang nasa content at dahil dito, ito ay "mapanlinlang." Ayon pa sa kaniya, "naberipika na" ang content ng isa sa mga fact-checker ng Meta sa Pilipinas.

3. Mga Patakaran at Pagsusumite ng Content ng Meta

I. Mga Patakaran sa Content ng Meta

a. Pamantayan ng Komunidad sa Maling Impormasyon

Content na Maaaring Tanggalin

13. Sa ilalim ng Pamantayan ng Komunidad sa Maling Impormasyon nito, tinatanggal ng Meta ang maling impormasyon kung saan posibleng direkta itong makapag-ambag sa peligro ng napipintong pisikal na pamiminsala o upang "mangialam sa paggana ng mga proseso sa pulitika [pangingialam sa botante o census]." Ang maling impormasyong may kinalaman sa pangingialam sa botante na dapat tanggalin ay tinutukoy batay sa isang listahan ng mga ipinagbabawal na maling impormasyon tungkol sa mga iskedyul o lokasyon ng botohan, at pagiging kwalipikado ng botante o kandidato.

Content na Kwalipikadong I-fact check ng Third Party

14. Para sa iba pang mga uri ng maling impormasyon sa pulitika, ipinahayag ng Meta na hindi ito nakapokus sa pagtanggal kundi sa "pagbawas sa paglaganap nito o paggawa ng kapaligirang bumubuo ng produktibong diyalogo." Sa Estados Unidos, itinigil na ng Meta ang programa nito ng pag-fact check ng third party at tinutugunan na nito ngayon ang mapanlinlang na content sa pamamagitan ng Mga Tala sa Komunidad. Sa labas ng Estados Unidos, isinasagawa pa rin ang pag-fact check. Para sa layuning ito, nakikipagtulungan ang Meta sa mga third-party na organisasyon ng pag-fact

check na sertipikado sa pamamagitan ng International Fact-Checking Network (o European Fact-Checking Standards Network sa Europe) na walang pinapanigang partido upang suriin at i-rate kung gaano katumpak ang pinaka-viral na content sa mga platform ng Meta.

15. Puwedeng suriin at i-rate ng mga fact-checker ang katumpakan ng mga pampublikong Facebook, Instagram, at Threads post, kabilang na rito ang mga ad, artikulo, larawan, video, Reel, post na audio at text lang. Ang mga opsyon sa pag-rate ay Hindi Totoo, Binago, Bahagyang Hindi Totoo, Kulang ng Konteksto, Satire at Totoo.

16. Pinagpapasyahan ng mga fact-checker kung anong content ang susuriin, at tinatapos nila ang kanilang pag-fact check nang independiyente sa Meta. Maaari nilang tukuyin ang content sa kanilang sariling inisyatiba o maaari silang pumili mula sa pila ng mga referral ng Meta para sa posibleng maling impormasyon. Nire-refer ng Meta ang content batay sa iba't ibang senyales, kabilang na rito ang "kung paano tumutugon ang mga tao," kung fina-flag ng mga user ang isang piraso ng content bilang "hindi totoong impormasyon," o kapag "nagpahayag ng pagtatwa" sa pagiging totoo nito ang mga komento sa isang post.

17. Isinasapriyoridad ng mga fact-checker ang viral na hindi totoong impormasyon at nabeberipika bilang hindi totoong claim na "napapanahon, trending at may malaking epekto" sa kinauukulang bansa at wika ng mga ito. Kabilang sa content na hindi kwalipikado para sa pag-fact-check ang "content na walang nebeberipikang claim," "opinyon at pananalita ng mga pulitiko," at "media na ginawa o in-edit sa paraang digital na naglalaman ng isa sa mga AI transparency label o watermark ng Meta batay sa pagiging totoo nito." Gayunpaman, kapag naglalaman ang minanipulang media ng hindi totoong claim na hiwalay sa paggamit ng media na ginawa o in-edit sa paraang digital, maaari pa ring i-rate ng mga fact-checker ang post. Ang na-fact check na content na ni-rate bilang Hindi Totoo, Binago, o Bahagyang Hindi Totoo ay maaaring i-demote, hindi irekomenda at tanggihan para sa mga ad.

Content na Kwalipikado para sa Paglalagay ng Label Bilang Minanipulang Media

18. Sa ilalim ng mga panuntunan sa minanipulang media para sa patakaran sa Maling Impormasyon, ang mapanlinlang na content na "ginawa o binago sa paraang digital" ngunit "hindi naman lumalabag sa ibang Mga Pamantayan

ng Komunidad” ay maaaring makatanggap ng pang-impormasyong label “sa harap ng content” o matanggihan kung isusumite bilang advertisement. Inilalapat ito sa isang “photorealistic na image o video, o realistikong tunog ng audio, na ginawa o binago sa paraang digital at nakalilikha ng partikular na malaking peligro na talagang makapanloko sa publiko sa isang usaping mahalaga sa publiko.” Hindi nagsasama ang Meta sa patakaran ng dagdag pang pamantayan para sa kung kailan “maaaring” ilapat ang label na ito.

19. Inaatasan ng Meta ang mga tao na isiwalat, sa pamamagitan ng “AI-disclosure tool” nito, tuwing magpo-post ang mga user ng “organic content na may photorealistic video o audio na may realistikong tunog na ginawa o binago sa paraang digital.” Ayon sa Meta, maaari itong magpataw ng mga kaparusahan kung hindi ito magagawa ng mga user.

20. Naglalapat ang Meta ng tatlong magkakaibang pang-impormasyong label para sa minanipulang media: ang label na “May Malaking Peligro,” ang label na “AI na Impormasyon” at ang label na “AI naMay Malaking Peligro.” Sa kaso ng Diumanong Audio na Tawag Upang Sirain ang Halalan sa Iraqi Kurdistan, ipinaalam ng Meta sa Board na upang makapagpataw ng label na May Malaking Peligro,dapat ang content ay (i) nakagagawa ng partikular na malaking peligrong talagang makakapanloko ng publiko sa isang usaping mahalaga sa publiko; at (ii) mayroong mga maaasahang indikasyong ginawa o binago ang content sa paraang digital. Ang label na AI na May Malaking Peligro ay katulad ng label na May Malaking Peligro, ngunit para ito sa content na may mga maaasahang indikasyong ginawa o binago ito gamit ang AI. Ang label na “AI na Impormasyon” ay para sa mga image, hindi video o audio, na ginawa ng AI na natukoy ng Meta sa pamamagitan ng “mga indikasyon ng AI image ayon sa pamantayan ng industriya o kapag isiniwalat ng mga tao na nag-upload sila ng AI-generated na content.”

21. Ang pagpapataw ng label na May Malaking Peligro ay hindi nagreresulta sa pag-demote sa content o sa pagkakatanggal nito mula sa mga rekomendasyon. Sa halip, kapag tinangkang i-reshare ng mga user ang isang post sa Facebook, Instagram o Threads na may label na May Malaking Peligro, makatatanggap sila ng pop-up na abisong mag-aalerto sa kanila na maaaring ginawa o binago ang content sa paraang digital.

b. Mga patakaran sa pag-demote

22. Inilalarawan sa Mga Alituntunin sa Pamamahagi ng Content ng Meta ang mga uri ng content na maaaring i-demote dahil sa pagiging "problematiko" o pagkakaroon ng "mababang kalidad" ng mga ito. Ibinabatay ng Meta ang pag-demote nito sa mga ganitong uri ng content sa "paninindigan nito sa mga kahalagahan ng Pagtugon sa Direktang Feedback ng mga Tao, Pagbibigay ng Insentibo sa mga Publisher Upang Mamuhunan Sila sa De-kalidad na Content at Pagbuo ng Mas Ligtas na Komunidad." Ayon sa Meta, binibigyang-daan nito ang mga tao na mag-share ng content "nang hindi nagagambala ng content na problematiko o may mababang kalidad."

23. Ayon sa Meta, kasama sa mga uri ng content na dine-demote nito ang na-fact check nang maling impormasyon pati na rin ang content na tinasa bilang "posibleng lumalabag" sa Mga Pamantayan ng Komunidad ng Meta. Kapag na-post sa Facebook ang content, tinatasa ito batay sa isang hanay ng mga classifier upang matukoy kung lumalabag ito sa isa sa Mga Pamantayan ng Komunidad ng platform. Nag-uulat ang mga classifier na ito ng degree ng "confidence" sa posibilidad ng paglabag na ginagamit ng kumpanya bilang senyales upang magsagawa ng ilang partikular na aksyon. Kapag mataas ang confidence score para sa partikular na content na iyon, ngunit hindi sapat na mataas upang mag-trigger ng agarang pagkakatanggal, maaaring i-demote ng Meta ang content. Kung mapag-alamang nilalabag ng content ang alinman sa mga saklaw na patakaran, tinatanggal ito. Maaaring i-adjust ng kumpanya ang confidence threshold na kinakailangan ng mga indibidwal na classifier upang magsagawa ng mga partikular na aksyon, gaya ng sa mga panahon ng krisis.

II. Mga Pagsusumite ng Meta

24. Ayon sa Meta, hindi posibleng masuri ng mga fact-checker ang lahat ng potensyal na maling impormasyon, kaya nagbigay ng tagubilin ang kumpanya sa mga fact-checker na isapriyoridad ang "mga viral at mahahalagang paksa na posibleng makapaminsala o kumalat agad-agad."

25. Sinabi ng Meta sa Board na umaasa ito sa dalawang pamamaraan upang maisapriyoridad ang content para sa pag-fact check sa mga kritikal na panahon. Una, puwedeng markahan ng Meta ang partikular na nakapilang content bilang agaran. Pangalawa, sa panahon ng mga pandaigdigang kaganapang may mataas na priyoridad gaya ng ilang partikular na krisis o halalan na isinasaalang-alang ng Meta bilang "Mga Trending na Kaganapan,"

fini-filter at ipinipila ng Meta ang kaugnay na content para sa pag-fact check. Gumagamit ang filter na ito ng listahan ng mahahalagang keyword na tinukoy ng mga lokal na espesyalista sa pagpapatakbo ng merkado na dalubhasa sa wika at konteksto gamit ang mga tinukoy na alituntunin. Ang content na tinukoy sa pamamagitan ng prosesong ito ay nakakatanggap ng label ng paksa na may angkop na pangalan ng kaganapan at isinasama sa nakatalaga at nakabatay sa oras na filter view ng Mga Trending ma Kaganapan sa loob ng tool para sa pag-fact check. Upang maging kwalipikado bilang Trending na Kaganapan, ang isang isyu ay dapat mahalaga (gaya ng isang krisis sa totoong mundo, sibikong debate, o may peligrong makapaminsala offline) at itinuturing na may malaking peligrong magpakalat ng maling impormasyon. Nananatili sa mga fact-checker ang pagpapasya kung pipili sila ng agaran o trending na content para sa pagsusuri.

26. Bukod pa sa pagpili ng content mula sa pila ng Meta, puwedeng tumukoy ang mga fact-checker ng content upang suriin nang independiyente sa mga platform ng Meta. Sinabi ng Meta sa Board na mayroong access ang mga fact-checker sa Meta Content Library, na isang web-based na tool na nagbibigay-daan sa mga may access dito upang magsagawa ng mga sistematikong paghahanap sa ilang content na naa-access ng publiko sa mga platform ng Meta. Puwedeng pumili ang mga fact-checker ng content na natuklasan sa Meta Content Library para sa pag-fact check. Sinabi ng Meta na “aktibo nitong iniimbitahan” ang mga fact-checker na gamitin ang tool na ito.

27. Hindi itinuring ng Meta ang pag-aresto ng ICC sa dating pangulo bilang trending na kaganapan at hindi nito sinertipika ang content bilang nangangailangan ng agarang pag-fact check. Sinabi ng kumpanya sa Board na hindi ito nakatanggap ng impormasyon tungkol sa mga trend ng maling impormasyon na humantong sa napipintong karahasan o pisikal na pamiminsala kaugnay ng midterm na halalan sa Pilipinas noong Mayo 2025. Gayunpaman, ibinahagi rin ng kumpanya na ang mga fact-checking partner nito ay naglabas ng “mga laganap na claim ng maling impormasyon gaya ng mga hindi totoong claim na may kinalaman sa pag-aresto sa dating Pangulong Rodrigo Duterte ng ICC at sa impeachment ng anak niyang si [Pangalawang Pangulong] Sara Duterte.”

28. Ipinaliwanag ng Meta na kapag na-rate na ng isang fact-checker ang isang content, sinasabi niya sa Meta kung nasaan ang hindi totoong impormasyon sa content (kung nasa caption ito, sa audio o video o bahagi lang nito, o kapag

parehong isinasaalang-alang ang media at caption). Pagkatapos, gumagamit ang Meta ng teknolohiya ng pagtutugma upang maipataw lang ang rating sa content na kaparehong-kapareho o halos kapareho para sa partikular na aspektong iyon (ibig sabihin, ang katulad na media, o ang katulad na media na dinagdagan ng katulad na caption). Kung ire-rate ng fact-checker ang video lang sa post, ipapataw ng Meta ang rating sa anumang post na naglalaman ng katulad o halos katulad na video. Gayunpaman, kung parehong ire-rate ng fact-checker ang video at caption na magkasamang isinasaalang-alang, ipapataw lang ang rating sa mga post na may katulad o halos katulad na video at caption.

29. Ginagamit ng Meta ang salitang "kaparehong-kapareho" na content upang ilarawan ang pag-share ng content na may "mga eksatong katulad na katangian" kagaya sa media na ni-rate ng mga fact-checker. Ang "halos kapareho" at "halos eksaktong katulad" na content ay may "kaunting pagkakaiba sa formatting o overlay text, ngunit may katulad na na-debunk na claim." Hindi awtomatikong nagpapataw ang Meta ng mga label nang malawakan kung hindi natutugunan ng content ang mga pamantayan para sa kaparehong-kapareho o halos kaparehong content. Dahil ito sa napag-alaman ng kumpanya na ang maliliit na pagkakaiba sa kung paano ipinapahayag ang isang claim ay puwedeng makaapekto sa kung totoo ito o hindi totoo. Para sa Meta, ang diskarteng ito ay "nakakatulong na pigilan ang di-wastong pagpapataw ng mga rating para sa pag-fact check."

30. Sa kasalukuyang kaso, napag-alaman ng Meta na hindi nakatugon ang kaso sa threshold para sa kaparehong-kapareho o halos kapareho na content upang maipataw ang katulad na label ng rating na naipataw ng mga fact-checker ng Meta para sa mga katulad na video na nagpapakita ng protesta sa Serbia bilang protestang pabor kay Duterte sa Netherlands. Halimbawa, ang ilang video na ni-rate ng mga fact-checker ay may ibang audio, dahil ang post na ito ay may awiting "Bayan Ko" sa background, o iba ang mga caption.

31. Ang content na ni-refer sa mga fact-checker para sa pagsusuri ay mawawala na sa pila ng mga referral pagkaraan ng pitong araw kung hindi pipiliin ng mga fact-checker ang content para sa pagsusuri. Ipinaliwanag ng Meta na ipinakikita sa time frame na ito ang trend na "nangyayari ang karamihan ng view sa loob ng mga unang araw ng lifecycle ng isang content." Maaari pa ring pumili ang mga fact-checker ng content na susuriin sa labas ng pila ng referral ng Meta, at mag-rate ng content "gaano man ito katagal na

nasa platform. Halimbawa, bagaman pangkalahatang nakapokus ang mga fact-checker sa mga napapanahong paksa, maaari pa ring lumutang ulit ang mas lumang content dahil sa mga bagong isyu."

32. Kaugnay ng mga panuntunan sa minanipulang media para sa patakaran nito sa Maling Impormasyon, hindi nagpataw ang Meta ng alinman sa mga label nito, kabilang na rito ang label na May Malaking Peligro, sa content ng kaso. Sinabi ng Meta sa Board na ito ay dahil hindi ito panloob na na-escalate. Noong inanalisa ito kasunod ng pagpili ng Board sa kaso, "mahigit 2 buwan na" ang content. Dahil sa "katagalan at pagiging hindi viral ng content," nagpasya ang Meta na hindi ipataw ang label sa minanipulang media sa puntong iyon. Binigyang-diin ng Meta na ang label na May Malaking Peligro ay isang patakaran para lang sa pag-escalate. Madalang itong ginagamit upang "tugunan ang content na ginawa sa paraang digital na may hatid na partikular na biglaang peligrong malinlang ang publiko tungkol sa isang mahalagang isyu sa kritikal na panahon." Ayon sa Meta, kasama rito ang "kapag na-post ang content nang malapit sa oras ng isang kritikal na kaganapan, gaya ng halalan, at walang sapat na panahon para sa ecosystem ng impormasyon o mga fact-checker na tugunan ang pinag-uusapang content." Dahil na-post ang content dalawang buwan bago ang midterm na halalan noong Mayo 2025 sa Pilipinas, para sa Meta, sapat na panahon ito para sa "ecosystem ng impormasyon o pangontrang pananalita upang maiwasto ang anumang maling impormasyon sa post (at dito, na-fact check ang mga katulad na bersyon ng video)." Lumalabas na hindi isinaalang-alang ng Meta ang alinman sa pagkakaaresto ni Duterte o sa mga kasunod na protesta bilang independiyenteng "mga kritikal na kaganapan."

33. Ang tanging pagpapatupad na isinagawa ng Meta sa content na ito ay pansamantalang pababain ang visibility ng post sa Facebook feed ng mga user na wala sa US, kabilang na ang mga nasa Pilipinas, batay sa mga senyales na nanghuhulang naglalaman ng maling impormasyon ang post.

34. Nagtanong ang Board ng mga bagay tungkol sa kung paano gumagana ang pag-fact check, kung paano itinatalaga ng Meta ang mga resource sa mga fact-checker, kung paano inihahambing ang magkakatulad na video tungkol sa pag-aresto kay Duterte sa ni-reshare na video, at kung isinaalang-alang ng Meta ang mga panuntunan sa minanipulang media para sa patakaran sa Maling impormasyon sa pag-moderate sa content ng kaso. Sinagot ng Meta ang lahat ng tanong.

4. Mga Komento ng Publiko

35. Nakatanggap ang Oversight Board ng anim na komento mula sa publiko na nakatugon sa mga tuntunin para sa pagsusumite. Lima sa mga komento ang isinumite mula sa Asia-Pacific and Oceania at isa mula sa Europe. Upang mabasa ang mga komento ng publiko na isinumite nang may pagpayag na mailathala ang mga ito, mag-click dito.

36. Nakapailalim ang mga pagsusumite sa mga sumusunod na tema: mga limitasyong kinahaharap ng mga fact-checker, lokal na pagkadalubhasa ng mga fact-checker, mga kakulangan sa kasalukuyang diskarte ng Meta sa pagtugon sa mga may koordinasyong kampanya ng pagpapakalat ng disimpormasyon, at mga account na responsable para sa pagbabahagi ng maling impormasyon sa Pilipinas.

37. Noong Hulyo 2025, bilang bahagi ng kasalukuyang pakikipag-ugnayan sa stakeholder, kinonsulta ng Board ang mga kinatawan ng mga organisasyon ng fact-checker, akademiko at iba pang eksperto sa maling impormasyon. Tinalakay ng roundtable kung paano gumagana ang pag-fact check, ang pag-fact check ng Meta sa labas ng Estados Unidos, ang dami ng posibleng mapanlinlang na content na kinahaharap ng mga fact-checker at ang kakayahan ng mga nagpapakalat ng disimpormasyon na gumawa ng mga pagbabago sa post upang hindi ito ma-detect, pati na rin ang peligro ng labis na pagpapatupad sa patakaran gamit ang teknolohiya ng pagtutugma upang ma-detect ang kaparehong-kapareho at halos kaparehong content.

5. Pag-aanalisa ng Oversight Board

38. Ang mga kampanya ng disimpormasyon ay naghahatid ng mga banta sa integridad ng impormasyon, tiwala ng publiko at sa mismong demokrasya. Pinili ng Board ang kasong ito upang suriin kung paano tinutugunan ng Meta ang hindi totoo o mapanlinlang na impormasyon sa mga platform nito, lalo na kapag na-share sa kasagsagan ng matinding tensyon sa pulitika at sa mga konteksto kung saan nakakaimpluwensya ang disimpormasyon sa debate ng publiko. Inanalisa ng Board ang desisyon ng Meta sa kasong ito batay sa mga patakaran sa content, bagay na pinahahalagahan at tungkulin para sa mga karapatang pantao ng Meta. Tinasa rin ng Board ang mga implikasyon ng kasong ito para sa mas malawak na diskarte ng Meta sa pangangasiwa ng content.

5.1 Pagsunod sa Mga Patakaran sa Content ng Meta

I. *Mga Panuntunan sa Content*

39. Napag-alaman ng Board na ang content ng kaso ay hindi nakakatugon sa mga pamatayan para sa pagtanggal sa ilalim ng pamantayan sa patakaran sa Maling Impormasyon. Hindi ito nagbibigay ng impormasyon tungkol sa alinman sa mga ipinagbabawal na kategorya sa ilalim ng linya ng patakarang ito, kabilang na rito ang mga lokasyon ng botohan o census, proseso ng pagboto, o pagiging kwalipikado ng botante o kandidato. Napag-alaman din ng Board na natutugunan ng content ng kaso ang mga pamantayan upang maging kwalipikadong ma-fact check sa ilalim ng patakaran ng Maling Impormasyon ng Meta, at sang-ayon ito sa Meta na isumite ang post para sa pag-fact check. Natutugunan ng content ng kaso ang mga pamantayan upang maging kwalipikadong malagyan ng label na minanipulang media sa ilalim ng patakaran sa Maling Impormasyon ng Meta. Samakatuwid, dapat nagpataw ang Meta ng label na "May Malaking Peligro" sa content.

A. Pamantayan ng Komunidad sa Maling Impormasyon at Third-Party na Pag-fact check

40. Tinutugunan ng content ng kaso ang mga pamantayan upang maging kwalipikado na ma-fact check sa ilalim ng patakaran sa Maling Impormasyon ng Meta. Ang post ay naglalaman ng nabeberipikang claim, walang kasamang pananalita mula sa mga pulitiko, at walang AI transparency label na dahilan upang maging hindi kwalipikado ang content para sa pag-fact check (ibig sabihin, ang mapanlinlang na katangian ng post ay nasa kung paano ito ginawa o binago sa paraang digital, at wala sa kine-claim ng post). Natutugunan din nito ang isa sa mga elemento upang maisapriyoridad ang content para sa pag-fact check — "napapanahon, trending at may malaking epekto" ito. Nai-share ang post ilang saglit pagkatapos maaresto ng dating Pangulong Duterte, kung kailan pinagtatalunan online at offline ang mga pangyayari tungkol sa kaniyang pagkakaaresto. Sa panahong na-post ang content ng kaso, sumiklab ang mga protesta na pabor at tutol sa pag-aresto kay Duterte. Ang mga post na katulad sa content ng kaso na nagke-claim na nagaganap ang mga protestang pabor kay Duterte ay nag-viral sa social media sa panahong iyon. Iniulat ng isa sa mga fact-checker ng Meta na umani ang mga katulad na video ng mahigit tatlong milyong view. Dagdag pa rito, naganap ang pag-aresto laban sa mas malawak na konteksto ng matinding away pulitika sa pagitan ng Pangulong Marcos at Pangalawang Pangulong

Sara Duterte, sa isang bansa kung saan nakaiimpluwensiya sa debate sa pulitika ang disimpormasyon.

41. Umaksyon dapat ang Meta upang isapriyoridad ang nasabing content sa pilang ibinibigay nito sa mga fact-checker. Minarkahan dapat ng kumpanya ang content ng kaso bilang agaran noong ipinila nito ang content para sa pag-fact check. Itinalaga rin dapat ng Meta ang posibleng maling impormasyon patungkol sa pag-aresto sa dating Pangulong Duterte bilang Trending na Kaganapan at aktibong tinukoy at inilabas ang katulad na content matapos i-rate ng mga fact-checker ng Meta bilang hindi totoo ang mga katulad na video. Ayon sa sariling mga pamantayan ng Meta, mahalaga ang isyu para sa "sibikong debate" at nagkaroon ng "malaking peligrong may kumalat na maling impormasyon." Lumalabas mula sa impormasyong ibinigay sa Board na naging pabaya ang Meta sa bagay na ito.

42. Naging kwalipikado rin dapat ang content ng kaso bilang halos kapareho ng nakaraang na-fact check na content at nalagyan ng label na gaya nito. Natukoy ng mga fact-checker na nag-rate ng mga katulad na video ang pagiging hindi tumpak ng pagpapares ng footage ng protesta sa Serbia at audio ng mga taong sumisigaw ng "Duterte!" sa background. Ang mga ni-rate na video na iyon ay hindi naglalaman ng awiting "Bayan Ko" bukod pa sa pagsigaw. Dahil kasama sa content ng kaso rito ang awiting ito, maituturing ito bilang sapat na naiiba para sa teknolohiya ng pagtutugma ng Meta upang hindi ito tukuyin bilang kaparehong-kapareho o halos kapareho ng mga katulad na video na dati nang ni-rate bilang Hindi Totoo.

43. Para sa Board, ang mga isinagawang pagbabago ay nagkaroon ng "mga kaunting variation," at inulit ng content "ang na-debunk nang claim" gaya ng sa mga katulad na video na ni-rate bilang hindi totoo ng mga fact-checker ng Meta. Hindi binago ng mga kaunting pagkakaiba sa audio, text at caption ang mga pangunahing mapanlinlang na katangian ng content na ito. Kapag hindi nakakaapekto ang mga hindi halatang pagbabago sa mapanlinlang na katangian, dapat itong ituring na halos kapareho.

44. Kinikilala rin ng Board na ang mga automated system na kasalukuyang ginagamit ng Meta ay maaaring walang kakayahang gawin ang ganitong pagpapasya nang malawakan. Ang pagpapataw ng label sa content na may ibang audio nang walang anumang ebalwasyon ay posibleng humantong sa paglalagay ng label sa content na may audio na pumupuna o nagde-debunk

sa mapanlinlang na video. Nauunawaan ng mga stakeholder na kinonsulta ng Board na hindi mahuhulaan ng teknolohiya ng pagtutugma ang lahat ng posibleng paggamit at nuance ng pagpapahayag, at ang pagpapalawak nito, nang hindi nagde-develop ng iba pang tool upang tugunan ang maling impormasyon, ay maaaring humantong sa labis na pagpapatupad sa patakaran. Kasabay nito, hindi mabisa ang kasalukuyang diskarte ng Meta sa pagtugon sa mga aktor na nagti-tweak ng content sa mga hindi halatang paraan upang paglaruan ang sistema at makaiwas sa pagpapatupad ng patakaran. Dapat magtalaga ang Meta ng sapat na mga resource, automated o hindi, upang matiyak na mabisang naipapataw ang mga rating ng mga fact-checker.

B. Pamantayan ng Komunidad sa Maling Impormasyon at Minanipulang Media

45. Napag-alaman ng Board na ang content ng kaso ay bumubuo sa minanipulang media na nakatanggap dapat ng label. Batay sa mga katulad na pampublikong pag-fact check, isang “photorealistic na video” ang content na mukhang “ginawa o binago sa paraang digital.” Isinaalang-alang ng Board ang video footage ng protesta sa Serbia bilang pagtugon sa kinakailangan para sa pagiging “photorealistic” ayon sa patakaran sa Minanipulang Media. Ang pagpapares ng footage ng protesta sa Serbia at audio ng mga taong sumisigaw ng pangalang Duterte! at pagdaragdag ng text overlay na "Netherland" dito ay isang digital na alteration ng orihinal na footage. Dahil sa text overlay na “Netherland,” lalong nagulo ang kahulugan nang pinalabas nito na nangyari ang protesta sa Netherlands, kahit na sa katunayan, nangyari ito sa Serbia. Ang post ay “hindi rin lumalabag sa iba pang Mga Pamantayan sa Komunidad” na nanghihingi ng ibang pagtrato, gaya ng pagtanggal. Naghahatid ang content ng "partikular na malaking peligrong talagang maloko ang publiko sa isang usaping mahalaga sa publiko” dahil patungkol ito sa isang makabuluhang isyu sa pulitika ng Pilipinas. Kumalat ang mga katulad na video sa mga platform ng Meta, at ni-rate ang mga ito bilang Hindi Totoo ng mga fact-checker ng Meta sa Pilipinas. Ang bilang ng katulad na mapanlinlang na content ay nakadagdag sa peligrong talagang maloko ang publiko. Batay sa mga salik na ito, naipataw dapat ng Meta ang label nito na May Malaking Peligro.

46. Iginiit ng Meta na hindi ito nagpataw ng anumang label para sa minanipulang media dahil hindi na-escalate ang content sa mga panloob na

pangkat ng Meta, at na-post ito nang masyadong maaga pa bago ang halalan noong Mayo 2025, kaya maaari pang magkaroon ng sapat na pangontrang pananalita na magde-debunk sa hindi totoong claim. Dagdag pa rito, noong panahong pinili ng Board ang kaso, hindi isinaalang-alang ng Meta na idagdag ang label dahil sa katagalan at pagiging hindi viral ng post.

47. Hindi sang-ayon ang Board sa Meta. Una, ang mga pangangatwirang ito ay lumalabas na walang anumang batayan sa mga patakaran ng Meta na nakatuon sa publiko. Pangalawa, kahit gamitin pa ang mga panloob na pamantayang ito, iba ang kongklusyong nabuo ng Board.

48. Itinuring dapat ng Meta ang pag-aresto sa dating Pangulong Duterte bilang kritikal na kaganapan kung saan kinakailangang i-escalate ang mga may kaugnayang post para sa posibleng pagpataw ng label na May Malaking Peligro. Sa ilalim ng sariling pamantayan ng Meta, ang panahon ng pag-aresto kay Duterte ay kumatawan sa isang tensyonadong sandali sa pulitika sa Pilipinas, na umani ng kritikal na pampublikong diskurso tungkol sa mga pangyayari tungkol sa pag-aresto at pagprotesta bilang suporta at laban sa dating pangulo sa Pilipinas at sa ibang bansa. Ang mga stakeholder na kinonsulta ng Board ay nagbigay-diin sa mabilis na paraan kung paano naglipana sa mga platform ng Meta ang mga naratibo patungkol sa pag-aresto kay Duterte. Ayon sa Meta, na-fact check ang mga katulad na bersyon ng video. Nangyari din ang pag-aresto sa kasagsagan ng matagalang away pulitika sa pagitan ng dalawang pinakamatataas na opisyal ng pamahalaan ng bansa, kung saan ipinaparatang ng Pangalawang Pangulong Sara Duterte na dinukot ang kaniyang ama.

49. Batay sa mga aksyon nito rito, lumalabas na nilimitahan ng Meta ang interpretasyon nito tungkol sa mga kritikal na kaganapan kung saan kwalipikadong lagyan ng label ang minanipulang media sa mga halalan, nang hindi kasama ang iba pang mga krisis sa pulitika. Isinaalang-alang ng Board ang pag-aresto at ang kasunod nitong ekstradisyon sa ICC ng isang dating pinuno ng estado bilang kritikal na kaganapan. Dahil sa sitwasyon sa pulitika sa Pilipinas sa panahon ng pag-aresto sa dating pangulo, nagkamali ang Meta na itrato lang ang paparating na halalan noong Mayo 2025 bilang ang natatanging kritikal na kaganapang may kinalaman dito. Samakatuwid, sa pamamagitan ng mabisang paglilimita sa pagtatalaga ng kritikal na kaganapan sa mga halalan lang, binalewala ng Meta ang paggamit sa sarili nitong mga panuntunan upang matulungang limitahan ang pagkalat ng

minanipulang media sa mga sitwasyon ng matinding tensyon sa pulitika. Ang content na saklaw ng mga panuntunan sa minanipulang media ay mukhang lubos na mas kakaunti kaysa sa lumalabas na saklaw ng patakaran.

50. Kahit na ipagpalagay na ang kritikal na kaganapan sa kasong ito ay ang halalan noong Mayo 2025, hindi lang dapat umasa ang Meta sa kakayahan ng mga fact-checking partner nito upang palitawin at i-rate ang mga claim ng maling impormasyon. Ang mga fact-checker na kinonsulta ng Board pati na rin ang pagsusumite ng mga pampublikong komento ay lubos na nakapagbigay-diin sa malulubhang limitasyon sa resource na kinahaharap ng mga fact-checker sa paggawa ng kanilang trabaho (PC-31362, PressOne (Pilipinas); (PC-31357, Foundation for Media Alternatives (Pilipinas)). Ipinaliwanag ng iba pang mga fact-checker kung paanong hindi nila nare-rate ang lahat ng posibleng bersyon ng magkakatulad na video. Mula sa dami ng content, ang mga nagpapakalat ng disimpormasyon ay gumagawa ng mga hindi halatang pag-tweak sa content upang magmukhang mga bagong post ang mga ito kumpara sa content na dati nang na-fact check at dahil dito, natatakasan nila ang teknolohiya ng pagtutugma ng Meta para sa mga kaparehong-kapareho at halos kaparehong content (PC-31357, Foundation for Media Alternatives (Pilipinas); PC-31349, Rappler (Pilipinas)). Binigyang-diin ng mga stakeholder na posibleng matabunan ang mga makatotohanang claim dahil sa tuloy-tuloy na paglipana ng mga mapanlinlang na post na walang label sa mga platform ng Meta (PC-31358, European Fact-Checking Standards Network). Naitala ng isa sa mga fact-checker ng Meta na ang mga post na ito ay bumubuo sa bahagi ng isang mas malawak na organisadong kampanyang magpakalat ng disimpormasyon online (PC-31349, Rappler (Pilipinas)). Dahil sa mga pangyayaring ito, nakabuo dapat ang Meta ng kongklusyon na “hindi [nagkaroon] ng sapat na panahon upang matugunan ng ecosystem ng impormasyon o mga fact-checker ang pinag-uusapang content” na humantong sana sa pagsasapriyoridad sa ganitong uri ng content para sa pag-fact check.

51. Kapag ang mapanlinlang na impormasyon ay mukhang bumubuo ng bahagi ng mas malawak na sistemikong kampanya ng disimpormasyon upang impluwensyahan ang opinyon ng publiko tungkol sa isang partikular na isyu sa pulitika o lipunan sa ilalim ng mas matitinding tensyon, higit na kinakailangang tugunan ang mga viral na mapanlinlang na post, lalo na kapag mabilis na nagbabago-bago ang mga estratehiya ng pagpapalaganap upang makatakas sa pag-detect at pagsusuri. Sa mga kasong tulad nito, dapat

ipataw ng Meta ang lahat ng angkop na patakaran nito (kabilang na rito ang pag-fact check at pag-label) at mga kaugnay na tool.

II. *Aksyon sa Pagpapatupad*

52. Sa desisyon sa Binagong Video ng Pangulong Biden, ipinahayag ng Board ang alalahanin nito tungkol sa kagawian ng Meta na mag-demote ng content na nire-rate ng mga third-party na fact-checker bilang "hindi totoo" o "binago," nang hindi ipinapaalam sa mga user o nagbibigay ng mga mekanismo para sa pag-apela. Sinabi ng Board sa desisyong iyon na "ang pag-[d]emote ng content ay mayroong malalaking negatibong epekto sa kalayaan sa pagpapahayag. Dapat suriin ng Meta ang mga patakarang ito upang matiyak na malinaw nilang natutukoy kung bakit at kailan dine-demote ang content, at nabibigyan ang mga user ng access sa isang mabisang remedyo (Artikulo 2 ng ICCPR)." Kasabay nito, kinikilala ng Board na maaaring mayroong mga pagkakataon kung saan kinakailangan ang pag-demote bilang ang mas hindi gaanong mapanghimasok na hakbang kaysa sa pagtanggal ng content (tingnan ang halimbawang Mga Post na Sumusuporta sa mga Riot sa UK, Mga Alegasyon ng Krimen Batay sa Nasyonalidad, Make-up Video para sa Pag-aasawa ng Bata sa Iran).

53. Sa kasalukuyang kaso, ang aksyon ng Meta na pansamantalang ipakita ang content ng kaso nang mas mababa sa feed para sa mga user na wala sa US, kabilang na rito ang mga nasa Pilipinas, ay nakaayon sa kagawian nito ng pag-demote ng content na muntik nang lumabag sa patakaran. Dahil napag-alaman din ng Board na ang content ay halos kapareho ng content na dati nang ni-rate bilang hindi totoo, ide-demote din ito sa ilalim ng patakaran ng Meta na i-demote ang content na nire-rate ng mga third-party na fact-checker bilang hindi totoo. Gayunpaman, iginiit din ng Board ang mabigat nitong alalahanin hinggil sa kakulangan ng linaw sa mga patakaran sa pag-demote, mga pagkakataon sa pag-apela, at ng posibleng epekto sa pagpapahayag sa pulitika.

5.2 Pagsunod sa mga Tungkulin para sa mga Karapatang Pantao ng Meta

54. Napag-alaman ng Board na ang pagpapanatili ng content sa platform, na may label para sa minanipulang media na May Malaking Peligro, ayon sa iniaatas sa sariling mga patakaran ng Meta, ay nakaayon sana sa mga tungkulin para sa mga karapatang pantao ng Meta. Sang-ayon ang Board na ang pag-refer ng Meta sa content para sa pag-fact check ay nakaayon sa mga

tungkulin nito para sa mga karapatang pantao, at sa kasong ito, ang pagpapabuti sa mga tool na available sa mga third-party na fact-checker upang mabigyang-daan ang kanilang pagsusuri ay isang paraan upang matupad ang mga tungkuling iyon.

Kalayaan sa Pagpapahayag (Artikulo 19 ICCPR)

55. Ang Artikulo 19 ng ICCPR ay naglalaan ng malawak na proteksyon sa pagpapahayag, kabilang na rito ang diskurso sa pulitika (Pangkalahatang Komento Blg. 34, talata 11-12). Naglalaan ito ng “partikular na matinding” proteksyon para sa “pampublikong debate patungkol sa mga pampublikong personalidad sa mundo ng pulitika at mga pampublikong institusyon” (Pangkalahatang Komento Blg. 34, talata 38 20; tingnan din ang Pangkalahatang Komento Blg. 25, talata 12 at 25). Binigyang-diin ng UN Human Rights Committee na mahalaga ang kalayaan sa pagpapahayag para sa pagsasagawa ng mga gawaing pampubliko at ang mabisang paggamit ng karapatang bumoto (Pangkalahatan Komento Blg. 34, talata 20; tingnan din ang Pangkalahatang Komento Blg. 25, talata 12 at 25).

56. Hindi puwedeng ang pagiging hindi totoo lang ang tanging batayan upang tanggalin ang pananalita sa ilalim ng internasyonal na batas para sa mga karapatang pantao (UN Report of the Secretary-General sa pagkontra sa disimpormasyon, A/77/287, talata 13). Puwede lang itong paghigpitan kung makakapasa ito sa three-part test ng legalidad, pagkalehitimo, at pangangailangan at kaakmaan. Noong 2024, maraming UN Special Rapporteur at UN Working Group ang gumawa ng pinagkaisang pahayag tungkol sa pagpapatibay sa demokrasya at mga karapatang pantao sa mga halalan sa buong mundo: “Dapat suriin at gawing transparent ng mga kumpanya ng social media ang kanilang mga proseso tungkol sa pag-moderate at mga algorithm ng content upang matiyak na hindi sila tumutulong sa pag-censor sa pagtutol at sa pagsulong ng maling impormasyon. Ang mga kumpanya ng teknolohiya ay dapat magsagawa ng due diligence o angkop na pagsisikap, mamuhunan sa pag-fact check, at pag-unawa sa mga lokal na wika at mga lokal na konteksto sa kanilang mga patakaran sa pag-moderate ng content.”

57. Kapag nagpataw ang isang estado ng mga restriksyon sa pagpapahayag, dapat nilang matugunan ang mga kinakailangan sa legalidad, lehitimong layon, at pangangailangan at kaakmaan (Artikulo 19, talata 3, ICCPR).

Kadalasang tinutukoy ang mga kinakailangang ito bilang “pagsusuri na may tatlong bahagi (three-part test).” Ginagamit ng Board ang balangkas na ito upang bigyang-kahulugan ang mga tungkulin para sa mga karapatang pantao ng Meta kaugnay ng Mga Panggabay na Prinsipyo ng UN sa Negosyo at mga Karapatang Pantao, na pinapanindigan ng Meta sa Pangkumpanyang Patakaran sa mga Karapatang Pantao nito. Ginagawa ito pareho ng Board kaugnay ng sinusuring desisyon sa indibidwal na content at sa sinasabi nito tungkol sa mas malawak na diskarte ng Meta sa pangangasiwa ng content. Ayon sa sinabi ng UN Special Rapporteur sa kalayaan sa pagpapahayag, bagaman “walang obligasyon ang mga kumpanya sa mga Pamahalaan, dahil sa epekto ng mga ito, kinakailangan nilang tasahin ang parehong uri ng mga tanong tungkol sa pagprotekta sa karapatan ng mga user nila sa kalayaan sa pagpapahayag” (A/74/486, talata 41).

I. *Legalidad (Pagiging Malinaw at Accessible ng mga Panuntunan)*

58. Ang prinsipyo ng legalidad ay nag-aatas na maging accessible at malinaw ang mga panuntunang naglilimita sa pagpapahayag, na ginawa nang may may sapat na katumpakan upang mabigyang-daan ang isang indibidwal na maglagay ng regulasyon sa kanilang asal nang naaayon (Pangkalahatang Komento Blg. 34, talata 25). Bukod pa rito, ang mga panuntunang ito ay “hindi maaaring magpataw ng walang kontrol na paghuhusga hinggil sa paghihigpit sa kalayaan sa pagpapahayag sa mga napapailalim sa pagpapatupad [sa mga ito]” at dapat “magbigay ng sapat na patnubay sa mga napapailalim sa pagpapatupad sa mga ito upang mabigyang-daan sila na matiyak kung anong mga uri ng pagpapahayag ang wastong pinaghihigpitan at kung ano ang hindi” (ibid). Ayon sa UN Special Rapporteur sa kalayaan sa pagpapahayag, kapag ipinataw sa pamamahala ng mga pribadong aktor sa online na pananalita, dapat ay malinaw at partikular ang mga panuntunan (A/HRC/38/35, talata 46). Ang mga taong gumagamit ng mga platform ng Meta ay dapat magkaroon ng access at pag-unawa sa mga panuntunan, at ang mga tagasuri ng content ay dapat magkaroon ng malinaw na patnubay hinggil sa pagpapatupad sa mga ito.

59. Malinaw ang Pamantayan ng Komunidad sa Maling Impormasyon at mga kaugnay na patakaran ng Meta ayon sa pagkakapataw ng mga ito sa content sa kasong ito, ngunit dapat pang mapabuti ang mga ito. Ang pananalitang nakatuon sa publiko ng patakaran sa Maling Impormasyon ay malinaw na nag-aabiso sa mga user tungkol sa mga angkop na panuntunan at sa iba-ibang kahihinatnan ng pag-post ng maling impormasyon at minanipulang media sa mga platform ng Meta (ibig sabihin, tinatanggal, fina-fact check ang

content, o nagpapataw rito ng label para sa minanipulang media). Samantala, ang mga uri ng content na kwalipikado para sa pag-fact check, mga pamantayan sa pagtatalaga ng mga fact-checker, at mga patakaran sa pag-demote ay makikita sa iba-ibang page ng Transparency Center.

60. Sa desisyon sa Diumanong Audio na Tawag Upang Sirain ang Halalan sa Iraqi Kurdistan, sinabi ng Board na dapat isaalang-alang ng Meta ang “pagsasama ng impormasyon tungkol sa lahat ng iba-ibang label para sa minanipulang media sa isang page sa Transparency Center upang madaling malaman ng mga user ang tungkol sa mga ito,” at iginigiit nito ang patnubay na ito rito.

61. Dagdag pa rito, nag-aalala ang Board na ang interpretasyon ng Meta sa mga panuntunan nito sa Minanipulang Media para sa mga paglalagay ng AI label ay magreresulta sa mas kaunting saklaw kaysa sa ipinahihiwatig sa pampublikong patakaran. Dapat ilarawan ng Meta ang iba't ibang uri ng AI label na ipinapataw nito, ang mga pamantayan kung paano ipataw ang mga ito at ang mga kahihinatnan ng mga ito. Sa kasalukuyan, ang pinakadetalyadong paglalarawan sa paggamit ng Meta sa tatlong AI label ay nasa kamakailang desisyon sa Diumanong Audio na Tawag Upang Sirain ang Halalan sa Iraqi Kurdistan ng Board. Ang mga panuntunang nakatuon sa publiko ay dapat sumasalamin sa panloob na panuntunan ng Meta na kinakailangan ng kritikal na kaganapan upang makapagpataw ng label na May Malaking Peligro sa isang piraso ng content, pati na sa kung ano ang kwalipikado bilang kritikal na kaganapan.

II. *Lehitimong Layon*

62. Ang anumang paghihigpit sa kalayaan sa pagpapahayag ay dapat ding maghabol sa isa o higit pa sa mga lehitimong layong nakalista sa ICCPR, kung saan kasama ang pagprotekta sa mga karapatan ng iba (Artikulo 19, talata 3, ICCPR). Sa mga nakaraang desisyon, pinanghawakan ng Board na ang pagprotekta sa karapatang lumahok sa mga gawaing pampubliko (Artikulo 25, ICCPR) ay isang lehitimong layon para sa patakaran sa Maling Impormasyon ng Meta (Diumanong Audio na Tawag Upang Sirain ang Halalan sa Iraqi Kurdistan, Binagong Video ng Pangulong Biden).

III. *Pangangailangan at Kaakmaan*

63. Sa ilalim ng Artikulo 19(3) ng ICCPR, ang pangangailangan at kaakmaan ay nag-aatas na ang mga paghihigpit sa pagpapahayag ay “angkop dapat upang makamit ang silbi nito bilang tagapagprotekta; ang mga ito ay instrumento dapat na pinakahindi nanghihimasok mula sa lahat ng maaaring magsilbi bilang tagapagprotekta; akma dapat ang mga ito sa interes na poprotektahan” (Pangkalahatang Komento Blg. 34, talata 34). Para sa layuning ito, puwedeng gawan muna ng kumpanya ng ebalwasyon ang mga tool na mayroon ito upang makamit ang lehitimong layon nang hindi nakagagambala sa pananalita; pangalawa, kung hindi ito posible, tukuyin ang tool na pinakahindi nanghihimasok sa pananalita; at ikatlo, tasahin kung mabisa ang hakbang na pinipili nito (A/74/486, talata 52).

64. Sa pagtatasa sa pangangailangan at kaakmaan ng mga hakbang na ito, pati na rin sa pagpapasya kung ano pang ibang hakbang ang kailangan, isinaalang-alang ng Board ang sumusunod: a) na na-post ang content sa gitna ng matinding tensyon sa pulitika sa Pilipinas, na may mga nagbabago-bagong naratibong kumakalat tungkol sa mga pangyayari kaugnay ng pag-aresto sa dating Pangulong Duterte; b) madalas na paggamit ng social media sa bansa; c) isang kasaysayan ng disimpormasyon at maling impormasyon sa bansa, partikular na sa mga pangyayari bago ganapin at habang nagaganap ang mga halalan; d) ang pangkalahatang watak-watak na katangian ng kapaligiran ng pulitika at media; e) ang tuloy-tuloy na pagsasagawang ilegal ang media ng balita at mga mamamahayag pati na rin ang pagbaba ng tiwala ng publiko sa mga institusyon ng media; f) ang posibilidad na binago ang media na nakalagay sa post, gaya ng nakasaad sa mga katulad na post na ni-rate bilang Hindi Totoo ng lahat ng tatlo sa mga fact-checker ng Meta sa bansa; at g) ang posibilidad na malinlang at makaimpluwensya sa publiko ang media na binago sa paraang digital tungkol sa isyung nagpaigting sa pagkakawatak-watak sa larangan ng pulitika sa bansa.

65. Nauunawaan ng Board ang mga aksyon ng Meta kaugnay ng content ng kaso: pagsusumite sa content ng kaso para sa pag-fact check, una, ilang saglit pagkatapos i-post ang content, at muli, nang piliin ng Board ang kasong ito; at pansamantalang pagpapakita sa content nang mas mababa sa Facebook feed ng mga user na wala sa US, kabilang na sa Pilipinas, batay sa mga senyales na nanghuhulang naglalaman ng maling impormasyon ang content.

66. Bagaman nagsagawa ng ilang aksyon ang Meta upang mapangasiwaan ang mga fact-checker na nagsusuri sa content, dapat ay nakapagsagawa pa ito ng mga karagdagang hakbang alinsunod sa mga patakarang nito. Halimbawa, na-tag dapat ng kumpanya ang content ng kaso bilang agaran noong ipinila nila ito para sa pag-fact check, at itinuring ang pag-aresto bilang Trending na Kaganapan na may sibikong debate. Dapat ay aktibo ring natukoy at pinalitaw ng Meta ang content na tulad nito, dahil na-flag ng mga fact-checker ng Meta ang mga claim para sa laganap na maling impormasyon tungkol sa pag-aresto sa dating pangulong Duterte, at na-rate ang katulad na content bilang Hindi Totoo ng mga fact-checker ng Meta.

67. Dahil sa katangian ng disimpormasyon sa pulitika sa mga konteksto gaya ng Pilipinas na inilalarawan sa itaas at sa mabilis nitong pagkalat, dapat sumubok ang Meta ng mga paraan upang matukoy at matugunan nang mas mahusay ang maling impormasyon sa pamamagitan ng lahat ng magagamit na diskarte. Tulad ng binigyang-diin sa pinagkaisang pahayag ng mga UN Special Rapporteur, dapat tumugon ang mga kumpanya sa maling impormasyon gamit ang mga kagawian sa pag-moderate ng content na nakabatay sa lokal na konteksto.

68. Dapat pagbutihin ng Meta ang mga mekanismo nito para sa pag-fact check upang masuri ang katulad na content sa dati nang na-fact check na content. Sa desisyon ng Binagong Video ng Pangulong Biden, nauunawaan ng Board na ang pagpataw ng label sa maliit na bahagi ng content "ay maaaring lumikha ng hindi totoong impresyon na likas na katiwa-tiwala ang content na walang label." Sa kasalukuyan, maaaring independiyenteng tumukoy ng content ang mga fact-checker para sa pag-fact check, na sa teorya ay nagbibigay-daan upang matugunan ang content na katulad ng mapanlinlang na impormasyong may label na. Gayunpaman, naipahayag ng mga stakeholder sa Board na ang CrowdTangle, ang transparency tool na pagmamay-ari ng Meta na inihinto na noong 2024, ay mahalaga sa prosesong ito. Bagaman magagamit ng mga fact-checker ang Meta Content Library, hindi tiyak ang aktuwal na paggamit ng mga fact-checker sa tool na ito (Tingnan ang mga Post na Kinabibilangan ng "From the River to the Sea (Mga Ilog hanggang sa Dagat)"). Naitala ng mga stakeholder na hindi pare-pareho ang kakayahan ng mga journalist sa pag-access sa tool, habang naitala naman ng iba na hindi madaling gamitin ang tool para sa pag-fact check, na nagpapahayag sa kawalan ng kakayahang maghanap ng text sa mga video, magsagawa ng reverse-image search at maghanap ng mga pampublikong

grupo ayon sa lokasyon. Batay sa input na ito, dapat ipagpatuloy ng Meta na aktibong makipag-ugnayan sa mga third-party na fact-checker upang matiyak na akmang-akma ang Meta Content Library para sa kanilang mga layunin.

69. Dapat ding sabayan ang Meta Content Library ng iba pang mga hakbang upang matugunan ang mga claim para sa mga laganap na maling impormasyon. Para sa layuning ito, inirerekomenda ng Board na mag-develop ang Meta ng mga tool sa pila nito para sa pag-fact check na nagbibigay-daan sa mga fact-checking partner upang mabilisang matukoy ang viral content na hindi kwalipikado bilang kaparehong-kapareho o halos kapareho, ngunit posibleng ma-rate din bilang mga claim na hindi totoo o mapanlinlang. Ilalabas nito ang mga claim para sa maling impormasyon katulad ng mga na-fact check na habang iniiwasan ang peligro ng labis na pagpapatupad sa patakaran mula sa pagpapalawak ng kahulugan ng "kaparehong-kapareho" at "halos kapareho" na content.

70. Dapat magsagawa ang Meta ng mga hakbang upang mapabuti pa ang pagtugon nito sa maling impormasyon, na sa ilang bahagi ay upang maiwasang maipasa ang bigat ng trabaho ng pagtugon sa mapanlinlang na impormasyon sa mga third-party na fact-checking partner. Sa kabila ng pahayag ng Meta na namumuhunan ito sa teknolohiya upang ma-detect nang mas mahusay ang "mga hindi halatang pinagkaiba ng content" na maaaring nagbabahagi ng mapanlinlang na impormasyon, ipinakikita sa kasalukuyang kaso ang pangangailangan para sa karagdagang pagpapabuti. Para sa umiiral na nitong programa ng pag-fact check, halimbawa, dapat tiyakin ng Meta na nasusuportahan at napaglalaanan ng resource ang mga partner nito upang magampanan nila ang mapanghamong trabahong iniaasa sa kanila ng Meta. Dapat ding tugunan ng Meta ang mga account na paulit-ulit na nagpapakalat ng maling impormasyon, kabilang na rito ang sa pamamagitan ng pagpapatupad sa linya ng patakaran nito sa Hindi Makatotohanang Asal tungkol sa may koordinasyong hindi makatotohanang asal.

71. Nag-aalala rin ang Board sa pagkabigo ng Meta na ipataw ang label na May Malaking Peligro sa content ng kaso, tulad ng iniaatas sa mga patakaran ng kumpanya. Napag-alaman ng Board na posible sanang nakatulong ito upang mapigilan ang higit pang pagkalat ng hindi totoo o mapanlinlang na impormasyon sa Pilipinas sa kritikal na panahon. Ang mga user na nakakita sa post ay naalertuhan sana na maaaring binago sa paraang digital ang

content, dahil sa audio ng mga taong paulit-ulit na sumisigaw ng “Duterte!” at sa awiting “Bayan Ko” sa background na ipinares sa video footage ng protesta sa Serbia. Mapapansin na ang epekto ng label na May Malaking Peligro ay upang magpakita ng pop-up sa mga user na gustong mag-reshare sa post, na maaaring binago sa paraang digital ang post na balak nilang i-share. Bagaman nagbibigay-impormasyon ang label, nagdudulot ang abisong ito ng friction na puwedeng makatulong upang mabawasan ang pagkalat ng mga mapanlinlang na post tungkol sa mga kasalukuyang kaganapan kapag posibleng umangat ang ganitong uri ng content. Katulad nito, sa Diumanong Audio na Tawag Upang Sirain ang Halalan sa Iraqi Kurdistan, nagka-isyu ang Board sa ginagawang pagpili ng Meta sa kung saan lang magpapataw ng label na May Malaking Peligro bago ang isang watak-watak na halalan sa Iraqi Kurdistan, na lalo pang nagbibigay-diin sa pagkalitong puwedeng maidulot ng hindi patas na paglalagay ng label sa mapanlinlang na content. Itinala ng Board ang pag-uulat tungkol sa naobserbahan nitong hindi pagkakapare-pareho sa pagpataw ng mga AI label sa content, kahit na may mga senyales na ng pag-generate ng AI na nakaayon sa pamantayan ng industriya. Habang sumisikat sa social media ang AI-generated na content, napakahalagang mapabuti pa ng Meta ang pagpapatupad sa patakaran nito sa ganitong larangan.

72. Sa kasong ito, malubhang nag-aalala ang Board na hindi nagpataw ang Meta ng label na May Malaking Peligro. Ginawa dapat ito ng Meta rito, kasabay ng programa ng pag-fact check nito sa pamamagitan ng paggamit sa iba pang mga nakalaang patakaran.

6. Ang Desisyon ng Oversight Board

60. Pinapanindigan ng Board ang desisyon ng Meta na hindi tanggalin ang content.

7. Mga Rekomendasyon

Patakaran sa Content

1. Upang mabigyang-impormasyon nang mas mabuti ang mga user kung paano ipinatutupad ang patakaran sa minanipulang media ayon sa Pamantayan ng Komunidad sa Maling Impormasyon, dapat ipaliwanag ng Meta ang iba-ibang pang-impormasyong label na ginagamit ng Meta para sa minanipulang media, kabilang na rito ang pagpapataw ng label na May Malaking Peligro kaugnay ng isang kritikal na kaganapan, at kung ano ang itinuturing bilang kritikal na kaganapan.

Isasaalang-alang ng Board na naipatupad na ang rekomendasyong ito kapag in-update na ng Meta ang mga salitang nakalagay sa Pamantayan ng Komunidad sa Maling Impormasyon upang maipakita ang pagbabago.

Pagpapatupad

2. Upang mabigyang-daan ang mga third-party na fact-checker upang mabisang matugunan ang mga pattern ng maling impormasyon, dapat bumuo ang Meta ng hiwalay na pila sa loob ng interface ng pag-fact check na kinabibilangan ng content na katulad, ngunit hindi kaparehong-kapareho o halos kapareho, sa content na na-fact check na sa isang partikular na market.

Isasaalang-alang ng Board na naipatupad na ang rekomendasyong ito kapag nakapagbahagi na ang Meta ng impormasyon sa Board na nagdedetalye sa bago nitong feature ng interface at kung paano nito binibigyang-daan ang mga fact-checker na magsama ng bago at katulad na content sa umiiral nang mga pag-fact check.

***Tala ukol sa Pamamaraan:**

- Ang mga desisyon ng Oversight Board ay ginagawa ng mga panel na binubuo ng limang Miyembro at inaaprubahan ayon sa pagboto ng mayorya ng buong Board. Ang mga desisyon ng Board ay hindi kailangang kumakatawan sa mga pananaw ng lahat ng Miyembro.

- Sa ilalim ng Karta nito, maaaring suriin ng Oversight Board ang mga apela mula sa mga user na may content na tinanggal ng Meta, mga apela mula sa mga user na nag-ulat ng content na hindi tinanggal ng Meta, at ang mga desisyong ni-refer ng Meta rito (Artikulo 2, Seksyon 1 ng Karta). Ang Board ay mayroong awtoridad na may legal na bisa upang panindigan o tutulan ang mga desisyon sa content ng Meta (Artikulo 3, Seksyon 5 ng Karta; Artikulo 4 ng Karta). Maaaring mag-isyu ang Board ng mga rekomendasyong walang legal na bisa na kinakailangang tugunan ng Meta (Artikulo 3, Seksyon 4; Artikulo 4 ng Karta). Kung mangangako ang Meta na kikilos ito ayon sa mga rekomendasyon, susubaybayan ng Board ang kanilang pagpapatupad.

- Para sa desisyong ito sa kaso, kinomisyon ang isang independiyenteng pananaliksik sa ngalan ng Board. Tinulungan ang Board ng Duco Advisors, na isang tagapayong kumpanyang nakapokus sa interseksyon ng geopolitics, pagtitiwala at kaligtasan, at teknolohiya.* Ang kaalaman mula sa mga eksperto sa larangan ng lingguwistika ay ibinigay ng Lionbridge Technologies, LLC, na may mga espesyalistang matatas sa mahigit 350 wika at nagtatrabaho mula sa 5,000 lungsod sa buong mundo.*