

## Fake newsy, trolle, boty i cyborgi w mediach społecznościowych

Adrian Grycuk

### Fake news, internet trolls, bots, and cyborgs in social media

This descriptive article looks at the role of internet trolls and their sockpuppets, social bots, their groups (social botnets), and cyborgs in the processes of creating and spreading falsehoods on social media platforms. It begins with an overview of these actors' activities, which are exacerbated by the ranking algorithms, social homophily, echo chambers, and filter bubbles. Next, the spread of fake news in social media is explored in more detail. The final section overviews the steps taken in 2016–2020 by technology companies to curb the volume and the users' exposure to fake news.

**Słowa kluczowe:** fake newsy, media społecznościowe, trolle internetowe, boty społecznościowe, cyborgi, dezinformacja w internecie, fałszywe konta w mediach społecznościowych

**Keywords:** fake news, social media, internet trolls, social bots, cyborgs, internet disinformation, fake social media accounts

### Wstęp

Rozpowszechnianie nieprawdziwych treści jest od setek lat stosowane jako narzędzie osiągnięcia celów politycznych i militarnych. Jednak pojawienie się internetu całkowicie zmieniło skalę i dynamikę takich działań. Cechy mediów społecznościowych oraz fakt, że są popularnym źródłem informacji, spowodowały, że nieprawdziwe lub częściowo nieprawdziwe treści, określane angielskim wyrażeniem *fake news* (fałszywe wiadomości), mogą błyskawicznie docierać do bardzo dużej liczby osób. Obserwowany w ostatnich latach wzrost zainteresowania prowadzoną w sieci dezinformacją, której jednymi z najważniejszych narzędzi są właśnie fake newsy, jest następstwem wydarzeń kampanii zorganizowanych przed referendum w sprawie dalszego członkostwa Wielkiej Brytanii w Unii Europejskiej oraz przed wyborami prezydenckimi w Stanach Zjednoczonych w 2016 r. W internecie rozpowszechniano wtedy wiele nieprawdziwych treści, co zdaniem niektórych badaczy mogło wpłynąć na wyniki obu tych głosowań. Również wśród ogromnej liczby wiadomości

na temat trwającej pandemii COVID-19 część jest nieprawdziwa i może stwarzać zagrożenia dla zdrowia publicznego, co zaczęto określać mianem infodemii<sup>1</sup>.

Celem artykułu jest syntetyczne przedstawienie zjawiska fake newsów oraz roli, jaką w tworzeniu i rozpowszechnianiu nieprawdziwych treści w mediach społecznościowych, w szczególności w serwisach (na platformach) społecznościowych takich jak Facebook, Instagram, TikTok, Twitter czy YouTube, odgrywają trolle internetowe i boty społecznościowe. Zawiera on również krótki przegląd sposobów rozpowszechniania na tych platformach nieprawdziwych treści oraz opis najważniejszych działań, które spółki będące właścicielami najpopularniejszych serwisów podjęły w latach 2016–2020 w celu ograniczenia liczby i widoczności tego rodzaju fałszywych informacji.

<sup>1</sup> *Infodemic Management*, World Health Organization, <https://www.who.int/teams/risk-communication/infodemic-management> [dostęp: 26 listopada 2020 r.].

**Tabela 1. Cele i sposoby wykorzystywania fake newsów w polityce**

Obszar	Opis
Wybory i referenda	jako narzędzie marketingu politycznego, m.in. do: dyskredytowania przeciwników politycznych i kreowania ich negatywnego wizerunku, tworzenia fałszywego wrażenia o wielkości społecznego poparcia dla kandydatów, partii lub kwestii będących przedmiotem referendum, a także wywierania wpływu na decyzje wyborców o wzięciu przez nich udziału w głosowaniu*
Konflikty międzynarodowe	do wywoływania i eskalowania konfliktów międzynarodowych (wykorzystywanie informacji jako broni; ang. <i>weaponization of information</i> ); w czasie tzw. wojen hybrydowych służą prowadzeniu działań propagandowych przeciwko innym państwom, kompromitowaniu ich polityków, osłabianiu zaufania obywateli do instytucji państwowych, wzmacnianiu radykalizacji postaw i wywoływaniu niepokojów społecznych
Nastroje społeczne i opinia publiczna	do oddziaływania na nastroje społeczne i manipulowania opinią publiczną m.in. przez zakłócanie i wpływanie na kierunki dyskusji, odwracanie uwagi od niewygodnych tematów i podważanie wiarygodności ekspertów; do organizowania pozornie spontanicznych oddolnych akcji obywatelskich (ang. <i>astroturfing</i> ), w szczególności demonstracji i protestów ulicznych
Polaryzacja społeczna	do pogłębiania istniejących i tworzenia nowych podziałów społecznych za pomocą nieprawdziwych treści dotyczących kwestii wywołujących emocje, takich jak m.in. aborcja, pedofilia, imigranci, uchodźcy, mniejszości narodowe, etniczne i religijne (m.in. Żydzi i muzułmanie), elity, prawa kobiet, osoby homoseksualne, biseksualne i transpłciowe (ang. <i>lesbian, gay, bisexual transgender</i> , LGBT), zmiany klimatu, kwestie historyczne i dotyczące zdrowia (szczepienia, technologia 5G, pandemia COVID-19)
Sytuacje kryzysowe	do wywoływania lub zwiększania paniki i chaosu oraz utrudniania działań mających na celu ograniczenie negatywnych skutków sytuacji kryzysowych (ataki terrorystyczne, katastrofy, epidemie)

\* Najczęściej są to nieprawdziwe informacje o warunkach udziału, terminie oraz sposobie głosowania.

Źródło: opracowanie własne.

## Fake newsy

Nie opracowano dotąd jednej, powszechnie akceptowanej definicji fake newsa. Pojęcie to cechuje duża pojemność znaczeniowa, a wobec samego zjawiska stosuje się także inne nazwy. Fake newsem określa się zwykle nieprawdziwe lub częściowo nieprawdziwe i celowo wprowadzające w błąd treści (wiadomości), które są rozpowszechniane w internecie oraz w mediach tradycyjnych (prasa, radio i telewizja). Są to jednak nie tylko treści całkowicie zmyślane, teorie spiskowe czy informacje pseudonaukowe, ale także wiadomości prawdziwe, które jednak zostały wyrwane z kontekstu, nieprecyzyjnie opisane co do sytuacji, miejsca lub tożsamości osób, błędnie datowane (zwykle pojawia się sugestia, że dotyczą bieżących wydarzeń) bądź uzupełnione mylnymi lub stroniczymi wnioskami. Przyjmuje się, że fake newsów nie stanowią natomiast treści będące dowcipem, parodią i satyrą ani też takie, z którymi ktoś się nie zgadza i których prawdziwość przez użycie tego – mającego jednoznacznie negatywny wydźwięk – określenia próbuje kwestionować. Ta ostatnia sytuacja jest jednym z powodów, dla których niektórzy badacze i eksperci zamiast wyrażenia *fake news* preferują słowo „dezinformacja”, a Snopes, jeden z najstarszych anglojęzycznych serwisów internetowych sprawdzających prawdziwość treści (ang. *fact-checking*), woli je zastępować określeniem *junk news* (wiadomości śmieciowe)<sup>2</sup>. Z kolei Facebook, największy serwis społecznościowy na świecie, nazywa takie treści *false news* (fałszywymi informacjami)<sup>3</sup>.

<sup>2</sup> B. Huberman, *'Fake News': Why Snopes Prefers Not to Say It Anymore*, <https://www.snopes.com/2019/07/09/fake-news-why-snopes-prefers-not-to-say-it-anymore> [dostęp: 2 listopada 2020 r.].

<sup>3</sup> [https://www.facebook.com/communitystandards/integrity\\_authenticity](https://www.facebook.com/communitystandards/integrity_authenticity) [dostęp: 20 listopada 2020 r.].

Fake news może być tekstem, grafiką, nagraniem wideo (filmem) lub ich połączeniem. W przypadku grafik są to głównie fotomontaże, zarówno fotografii, jak i zrzutów ekranów (np. zawierających posty z serwisów społecznościowych). Z kolei zmanipulowane filmy dzieli się na dwie kategorie, które są określane w języku angielskim jako<sup>4</sup>:

- *shallowfakes* lub *cheapfakes*, czyli filmy zmodyfikowane za pomocą znanych od dawna narzędzi (jest to np. zmiana szybkości odtwarzania czy wycięcie fragmentów nagrania);
- *deepfakes*, czyli filmy, w których obraz i dźwięk zostały zmienione za pomocą narzędzi sztucznej inteligencji najczęściej w taki sposób, że pokazane w nich osoby wypowiadają słowa, których nigdy nie powiedziały, lub robią rzeczy, których nigdy nie zrobiły<sup>5</sup>.

Fake newsy są tworzone, a następnie rozpowszechniane:

- z powodów politycznych, dla realizacji celów przez polityków, partie polityczne, organizacje (w tym terrory-

<sup>4</sup> I. Sample, *What Are Deepfakes – and How Can You Spot Them?*, „The Guardian”, 13 January 2020, <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them> [dostęp: 22 grudnia 2020 r.].

<sup>5</sup> Do produkcji *deepfakes* wykorzystuje się tzw. uczenie głębokie (ang. *deep learning*, do czego nawiązuje pierwszy człon nazwy), w szczególności generatywne sieci przeciwstawne (ang. *generative adversarial networks*, GANs). Uczenie głębokie to jedna z dziedzin uczenia maszynowego, będącego z kolei jednym z obszarów sztucznej inteligencji. Media stworzone lub zmienione za pomocą sztucznej inteligencji określa się często angielskim zwrotem *synthetic media* (media syntetyczne). Zob. *Deepfakes, Shallowfakes, and Speech Synthesis: Tackling Audiovisual Manipulation*, European Parliamentary Research Centre, <https://sciencemediahub.eu/2019/12/04/deepfakes-shallowfakes-and-speech-synthesis-tackling-audiovisual-manipulation> [dostęp: 22 grudnia 2020 r.].

styczne), grupy interesu oraz rządy krajowe i państw obcych<sup>6</sup> (tabela 1);

- z powodów finansowych, z chęci uzyskania przez ich autorów przychodów z reklam wyświetlanych na stronach internetowych, na których opublikowali oni nieprawdziwe treści<sup>7</sup>, jako komercyjna usługa, a także do manipulowania kursami spółek notowanych na giełdzie;
- z powodów ideologicznych (posiadanych poglądów i przekonań);
- jako narzędzie nieuczciwej walki konkurencyjnej w sektorze prywatnym (tzw. czarny PR);
- dla rozrywki, zabawy, z chęci zwrócenia na siebie uwagi lub uzyskania rozgłosu.

Wiedza na temat motywów wymyślania i rozpowszechniania nieprawdziwych treści ma istotne znaczenie dla działań podejmowanych w celu ograniczenia tego zjawiska.

Fake newsy są rozpowszechniane ze zmiennym natężeniem w różnych mediach w różnych państwach. W zdecydowanej większości nie naruszają one prawa, chociaż do ich tworzenia mogą być wykorzystywane informacje lub materiały pozyskane nielegalnie (zwykle w wyniku ataków hakerskich). W mediach społecznościowych mogą być rozpowszechniane ze zhakowanych kont, zwykle należących do znanych osób, instytucji lub organizacji<sup>8</sup>.

Wiele osób nie odróżnia fake newsów od wiadomości prawdziwych<sup>9</sup>. Wpływ tego rodzaju wiadomości na odbiorców nie został jednak jeszcze w dostatecznym stopniu zbadywany<sup>10</sup>, a wiedza na temat ich skutków politycznych, społecznych, gospodarczych i finansowych jest wciąż mała. Najwięcej badań dotyczy nieprawdziwych treści w serwisie Twitter, co wynika m.in. z otwartości jego interfejsu programistycznego (ang. *application programming interface*, API) oraz dużej popularności wśród polityków i dziennikarzy. Wiadomo, że fake

newsy, zwłaszcza te będące mową nienawiści, mogą prowadzić do przemocy, morderstw i niepokojów społecznych<sup>11</sup>. Jednocześnie dla wielu internautów nieprawdziwe treści są tylko nową kategorią niepotrzebnie angażującego spamu. Możliwe do pozyskania dane dotyczące zasięgu fake newsów w niektórych serwisach nie dają odpowiedzi na pytanie, czy istnieją związki przyczynowo-skutkowe między zapoznaniem się z fałszywymi treściami a późniejszą zmianą zachowań czy podjętymi w danej sprawie działaniami. Wiadomo natomiast, że dementowanie nieprawdziwych wiadomości jest często mało skuteczne. Wiele osób z trudnością zmienia zdanie wyrobione w czasie zapoznawania się z takimi treściami po raz pierwszy<sup>12</sup>. Sprostowania nie docierają do wszystkich, którzy się z fake newsami zetknęli, a niektórych mogą wręcz utwierdzać w przekonaniu o ich prawdziwości<sup>13</sup>.

## Trolle internetowe i ich pacynki

Kluczową rolę w tworzeniu i rozpowszechnianiu fake newsów odgrywają tzw. trolle internetowe (dalej: trolle). W podstawowym znaczeniu tego słowa trolle to osoby, które swoimi kontrowersyjnymi lub napastliwymi wypowiedziami chcą sprowokować emocjonalne reakcje innych internautów. Zjawisko trollowania istnieje praktycznie od początku istnienia sieci<sup>14</sup>. Od kilku lat słowem „troll” zaczęto jednak także określać osoby dodające treści i wspierające określone narracje za wynagrodzeniem. Zwykle działają one na zlecenie wyspecjalizowanych podmiotów zajmujących się marketingiem politycznym i internetowym, nazywanych zwyczajowo farmami trolli (ang. *troll farms*)<sup>15</sup>. Są one ważnym graczem na rynku usług manipulacyjnych w mediach społecznościowych, któ-

<sup>6</sup> S. Bradshaw, P.N. Howard, *The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation*, Computational Propaganda Research Project, University of Oxford, Oxford 2019, <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf> [dostęp: 31 grudnia 2020 r.].

<sup>7</sup> Wykorzystując serwisy reklamowe takie jak Google AdSense.

<sup>8</sup> Zob. np. P. Domm, *False Rumor of Explosion at White House Causes Stock to Briefly Plunge; AP Confirms Its Tweeter Feed Was Hacked*, 23 April 2013, <https://www.cnbc.com/id/100646197> [dostęp: 31 grudnia 2020 r.].

<sup>9</sup> C. Silverman, J. Singer-Vine, *Most Americans Who See Fake News Believe It, New Survey Says*, <https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey> [dostęp: 20 grudnia 2020 r.].

<sup>10</sup> D.M.J. Lazer i in., *The Science of Fake News*, „Science” 2018, vol. 539(6380), s. 1095, <https://doi.org/10.1126/science.aao2998>; J.A. Tucker i in., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, 2018, s. 57–58, <https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf> [dostęp: 2 sierpnia 2020 r.]; E. Zhuravskaya, M. Petrova, R. Enikolopov, *Political Effects of the Internet and Social Media*, „Annual Review of Economics” 2020, vol. 12, s. 429, <https://doi.org/10.1146/annurev-economics-081919-050239>. Zdaniem niektórych badaczy ten wpływ może być jednak mniejszy, niż się powszechnie uważa. Zob. np. *A Russian Troll Farm May Not Have Been Very Good at Its Job*, <https://www.technologyreview.com/2019/11/25/131832/russia-disinformation-twitter-internet-research-agency-social-media-politics> [dostęp: 2 listopada 2020 r.].

<sup>11</sup> Zob. np.: M. Safi, *WhatsApp Murderers: India Struggles to Combat Crimes Linked to Messaging Service*, 3 July 2018, <https://www.theguardian.com/world/2018/jul/03/whatsapp-murders-india-struggles-to-combat-crimes-linked-to-messaging-service> [dostęp: 28 grudnia 2020 r.]; S. Stecklow, *Hatebook. Why Facebook Is Losing the War on Hate Speech in Myanmar*, 15 August 2018, <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate> [dostęp: 28 grudnia 2020 r.].

<sup>12</sup> M.S. Chan, C.R. Jones, K.H. Jamieson, D. Albarracin, *Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation*, „Psychological Science” 2017, vol. 28(11), <https://doi.org/10.1177/0956797617714579>. W 2017 r. Facebook, powołując się na badania naukowe, wycofał wprowadzone w poprzednim roku czerwone ikonki i krótkie ostrzegawcze komunikaty dotyczące treści zakwestionowanych przez zewnętrznych partnerów. Zaczęto za to zachęcać użytkowników do zapoznania się z podlinkowanymi artykułami pochodzącymi ze źródeł wysokiej jakości, <https://about.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation> [dostęp: 10 grudnia 2020 r.].

<sup>13</sup> D.J. Flynn, B. Nyhan, J. Reifler, *The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics*, „Political Psychology” 2017, vol. 38, s. 130–131, <https://doi.org/10.1111/pops.12394>.

<sup>14</sup> Podstawową strategią obronną użytkowników wobec trolli jest ignorowanie ich i niewchodzenie z nimi w dyskusje, co zwyczajowo nazywa się „niekarmieniem” trolli. Często też trolle są blokowane przez moderatorów i administratorów, a ich wkład jest usuwany.

<sup>15</sup> Najbardziej znaną farmą trolli jest rosyjska spółka Agentsvo Internet Issledovaniy (ros. Агентство интернет-исследований, ang. Internet Research Agency) z siedzibą w Petersburgu.

ry rozwinął się w ostatnich latach<sup>16</sup>. Funkcjonowanie farm, a także inne, luźniejsze formy koordynowania działalności płatnych trolli<sup>17</sup> pozwalają wywierać wpływ na kierunki dyskusji toczących się w internecie i uzyskiwać przez niektóre treści o wiele większe zasięgi (liczby odsłon) o wiele szybciej, niż gdyby rozpowszechniały się one w sposób organiczny (tj. w wyniku zainteresowania i dobrowolnych działań internautów).

Płatne trolle mogą świadczyć pracę zarówno w biurach farmy (z wykorzystaniem jej infrastruktury, zwykle w systemie zmianowym), jak i zdalnie. Otrzymują do wykonania codzienne cele ilościowe, takie jak liczba opublikowanych postów i dodanych komentarzy, często będące równocześnie podstawą do rozliczeń finansowych farm z ich klientami. Mogą być to jednak również cele jakościowe, np. liczba otrzymanych przez ich wkład polubień, komentarzy i udostępnień. Farmy produkują i przekazują swoim trollom grafikę i filmy potrzebne do realizacji zleczanych im zadań oraz koordynują ich współpracę.

Płatne trolle tworzą konta w mediach społecznościowych, na portalach, a także na stronach internetowych należących do mediów tradycyjnych, które dopuszczają możliwość komentowania publikowanych tam treści. Mogą one również dodawać treści z już istniejących kont należących do farmy. Bardzo ważnym obszarem ich aktywności są otwarte i zamknięte grupy dyskusyjne, które często same tworzą i animują. Do promowania swoich treści wykorzystują także zyskujące popularność hasztagi<sup>18</sup>. W odróżnieniu od opisanych w dalszej części artykułu botów społecznościowych, trolle są ludźmi, a więc mogą tworzyć rozbudowane wypowiedzi i w interesujący sposób argumentować, wywierając wpływ na działania innych użytkowników i przebieg dyskusji. Wiele z nich bardzo dobrze zna języki obce i z powodzeniem prowadzi działania w środowiskach internetowych innych państw.

Trolle tworzą i powszechnie wykorzystują dodatkowe, fałszywe tożsamości, czyli pacynki (ang. *sockpuppets*). Tak nazywane są konta należące do tego samego użytkownika niebędące na danej platformie jego podstawowym kontem. Mogą je tworzyć wszyscy użytkownicy (nie tylko trolle). Pacynki są

zaliczane do większej kategorii kont określanych jako fałszywe, fikcyjne lub – zwyczajowo – fejkowe (ang. *fake*), czyli należących do lub kontrolowanych przez kogoś innego niż osoby uwidocznione w ich opisach<sup>19</sup>. Pacynki trolli najczęściej umożliwiają publikowanie przez tę samą osobę wiele razy takiej samej lub podobnej treści i stymulowanie interakcji użytkowników z taką treścią. Są także wykorzystywane do manipulowania dyskusjami, m.in. poprzez robienie w nich tzw. sztucznego tłoku. Profile kont pacynek zawierają nieprawdziwe dane osobowe oraz fotografie, które najczęściej zostały pozyskane z sieci z naruszeniem prawa do ochrony wizerunku przedstawionych na nich osób i majątkowych praw autorskich<sup>20</sup>. Pacynki są również często zręcznie „hodowane”, tj. trolle przypisują im wybrane cechy, poglądy czy zainteresowania i uwiarygodniają je spójną historią ich aktywności. Aby zatrzeć związki pacynek z kontami głównymi, trolle często różnicują ich style wypowiedzi i sieci znajomych (to ostatnie ułatwia akceptowanie przez wielu użytkowników zaproszeń od nieznanym im osób). Duża liczba kontaktów zwiększa wiarygodność pacynek i zasięgi rozpowszechniania przez nie treści. Ponieważ jednak „hodowanie” kont jest czasochłonne, pacynkami mogą być również konta przejęte lub kupione wraz z ich historią i kontaktami.

Najciemniejszą stroną działalności trolli jest nękanie i przeprowadzanie ataków osobistych w celu zdyskredytowania, ośmieszenia, zastraszenia lub zmuszenia zaatakowanych osób do zrezygnowania z prowadzonych przez nie działań oraz do niepodejmowania takich działań (także przez inne osoby wiedzące o ataku) w przyszłości (tzw. efekt mrożący; ang. *chilling effect*). Publikowane podczas takich akcji treści są zwyczajowo określane jako „hejt” (ang. *hate*; nienawidzić). Trolle mogą wtedy również ujawniać dane osobowe, adresowe oraz tzw. dane wrażliwe zaatakowanych osób i ich bliskich (ang. *doxing*).

## Boty i cyborgi

Boty to aplikacje (oprogramowanie użytkowe) automatyzujące określone przez ich twórców czynności. Pierwotnie wykonywały one głównie proste, powtarzalne zadania o charakterze technicznym i porządkowym<sup>21</sup>. Obecnie, dzięki

<sup>19</sup> Mają one kluczowe znaczenie dla działań dezinformacyjnych w mediach społecznościowych, gdyż bez utworzenia konta korzystanie z nich jest niemożliwe. Precyzyjne zdefiniowanie fejkowego konta jest jednak trudne. Do tej kategorii nie są zaliczane jedyne konta należące do osób edytujących pod pseudonimami lub zmienionymi imionami i nazwiskami, co wynika najczęściej z chęci zachowania anonimowości i bardziej swobodnego uczestniczenia w dyskusjach.

<sup>20</sup> W celu utrudnienia ustalenia tożsamości przedstawionych na zdjęciach osób i źródeł ich pochodzenia (pierwotnego użycia) pozyskiwane z internetu fotografie są często modyfikowane za pomocą programów graficznych. Coraz częściej są również kreowane sztucznie, np. przy wykorzystaniu wspomnianych sieci GAN. Zob. np. R. Metz, *These People Do Not Exist. Why Websites Are Churning Out Fake Images of People (and Cats)*, <https://edition.cnn.com/2019/02/28/tech/ai-fake-faces/index.html> [dostęp: 22 grudnia 2020 r.].

<sup>21</sup> Dobrym przykładem takiego wykorzystywania botów pozostaje Wikipedia, w której przypadku tworzone i kontrolowane przez doświadczonych użytkowników boty m.in. dodają szablony, poprawiają linki, zmieniają nazwy kategorii i wprowadzają w hasłach zmiany typograficzne.

<sup>16</sup> A. Mierzyńska, *Fejkowe konta i hejt na zamówienie. Rynek usług manipulacyjnych kwitnie*, 12 kwietnia 2019, <https://oko.press/fejkowe-konta-i-hejt-na-zamowienie-rynek-uslug-manipulacyjnych-kwitnie-takze-w-polsce> [dostęp: 11 grudnia 2020 r.]; S. Bay, A. Dek, I. Dek, R. Fredheim, *Social Media Manipulation 2020*, NATO StratCOM COE, Riga 2020. Brakuje danych finansowych na ten temat, jednak uważa się, że największą część tego rynku stanowią zlecenia komercyjne polegające na kupowaniu sztucznej popularności i widoczności w mediach społecznościowych.

<sup>17</sup> Zjawiskiem tego typu o największej znanej skali są działania internetowych komentatorów w Chinach, wspierających swoimi wpisami narracje chińskiego rządu i Komunistycznej Partii Chin. Zob. np.: Y. Yang, *China's Communist Party Raises Army of Nationalist Trolls*, „Financial Times”, 30 December 2017, <https://www.ft.com/content/9ef9f592-e2bd-11e7-97e2-916d4fbac0da> [dostęp: 25 września 2020 r.]; G. King, J. Pan, M.E. Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument*, „American Political Science Review” 2017, vol. 111(3), s. 484, <https://doi.org/10.1017/s0003055417000144>.

<sup>18</sup> Oznaczenie dodawanych treści hasztagami niezgodnymi z intencjami ich twórców jest określane w języku angielskim jako *hashtag hijacking* (porywanie hasztagów). Takie działania prowadzą również boty.

zwiększającemu się wykorzystywaniu w informatyce sztucznej inteligencji, stają się coraz bardziej zaawansowane. Boty zaczęły być także powszechnie wykorzystywane w mediach społecznościowych. Te, które wykorzystując tworzone w tych mediach konta, naśladowują zachowania użytkowników, są nazywane botami społecznościowymi (ang. *social bots*; dalej: boty). Botami zaczęto również określać same zautomatyzowane konta<sup>22</sup>. W zależności od celu i sposobu wykorzystywania można je umownie podzielić na „dobre” i „złe”. Te pierwsze wykonują szereg typowych dla botów pożytecznych czynności (np. publikują przypomnienia lub aktualizacje na stronach). Działania tych drugich, opisywanych w niniejszym artykule, nie dodają wartości lub są uznawane za szkodliwe. Przykładem mogą być boty wykorzystywane do celów politycznych, nazywane czasem botami politycznymi (ang. *political bots*).

Boty służą do:

- szybkiego rozpowszechniania treści, w tym zwłaszcza nowo opublikowanych;
- tworzenia nowych treści;
- budowania wrażenia większej niż rzeczywista popularności osób, stron (profilu, fanpage'ów), postów, a także poparcia dla diskutowanych kwestii lub ocen bieżących wydarzeń czy też sprzeciwu wobec nich;
- wchodzenia w interakcje z użytkownikami.

Wkład większości botów to głównie polubienia, krótkie komentarze z hasztagami i udostępnienia. W serwisach społecznościowych są także tworzone grupy botów, czyli botnety społecznościowe (ang. *social botnets*)<sup>23</sup>. Ich działania są zwykle zarządzane przez jedną osobę. Działając w grupie, boty „współpracują” ze sobą, m.in. udostępniają nawzajem swoje posty i dodają polubienia, co zwiększa widoczność promowanych treści oraz postrzeganą przez użytkowników popularność narracji.

Podobnie jak trolle w przypadku wykorzystywania pacynek, operatorzy botów starają się upodobnić zautomatyzowane konta do kont zwykłych użytkowników. Na fakt, że konto zostało zautomatyzowane, może wskazywać m.in. to, że:

- zostało utworzone niedawno;
- wykazuje nienaturalnie dużą aktywność (liczba działań, która jest niezwykle trudna lub wręcz niemożliwa do wykonania w takim czasie przez człowieka);
- jest aktywne bardzo nieregularnie i w nietypowych porach dnia (np. nocami w godzinach od 1 do 6);
- jego wkład jest merytorycznie niespójny lub nie pasuje do opisu użytkownika;
- brakuje oryginalnego wkładu (rozpowszechnia wyłącznie treści innych użytkowników);
- widoczna jest niekompletność lub niespójność danych osobowych uwidocznionych na profilu (stronie użytkownika);
- brakuje wpisów dotyczących życia prywatnego (zainteresowania, rodzina, hobby).

<sup>22</sup> K.-C. Yang i in., *Arming the Public with Artificial Intelligence to Counter Social Bots*, „Human Behavior and Emerging Technologies” 2018, vol. 1(1), s. 48, <https://doi.org/10.1002/hbe2.115>.

<sup>23</sup> J. Zhang, R. Zhang, Y. Zhang, G. Yan, *The Rise of Social Botnets: Attacks and Countermeasures*, „IEEE Transactions on Dependable and Secure Computing” 2016, vol. 16(6), s. 1068, <https://doi.org/10.1109/tdsc.2016.2641441>.

Niektóre boty mogą odpowiadać na kierowane do nich pytania i wchodzić w dyskusje z użytkownikami, jednak nawet najbardziej zaawansowanym z nich sprawia trudność rozpoznanie w kierowanych do nich wypowiedziach aluzji, sarkazmu i humoru. Stąd ich odpowiedzi mogą być wtedy niespójne i nienaturalne.

Chociaż istnieją narzędzia pomagające identyfikować boty (głównie w serwisie Twitter), to nie są one stuprocentowo skuteczne. Dodatkowo operatorzy botów mogą automatyzować niektóre konta tylko na pewien czas lub uzupełniać ich działania swoimi. Wynika to m.in. z chęci zmniejszenia ryzyka zablokowania konta przez serwis i usunięcia jego wkładu czy też wejście przez bota w interakcje z użytkownikami, z którymi sam nie jest w stanie sobie poradzić. Takie hybrydowe konta, których aktywność jest połączeniem wkładu ludzi i botów, są nazywane cyborgami (ang. *cyborgs*)<sup>24</sup>.

Zautomatyzowane konta mogą być kontami zarówno nowo utworzonymi, jak i przejętymi lub kupionymi. Boty i ich usługi również stały się bowiem przedmiotem obrotu gospodarczego. Można kupić boty o różnym stopniu zaawansowania i z odmiennymi historiami działalności. Są one powszechnie używane przez farmy trolli, a w celu zwiększenia produktywności używa ich także wiele płatnych trolli. W ten sposób te same boty mogą być wykorzystywane w różnych państwach (w różnych wersjach językowych tego samego serwisu)<sup>25</sup>. W internecie można znaleźć również praktyczne informacje na temat samodzielnego budowania botów, przez co to narzędzie stało się dostępne także dla osób nieposiadających dużych umiejętności programistycznych.

Liczba i aktywność botów są różne w poszczególnych serwisach. Największy wpływ na to mają skuteczność ich właścicieli w identyfikowaniu i usuwaniu zautomatyzowanych kont (wraz z ich wkładem), które są sprzeczne z regulaminami serwisów, oraz dynamika bieżących wydarzeń (zwłaszcza politycznych). Według szacunków Facebooka liczba fałszywych kont (w tej kategorii znajdują się konta zautomatyzowane) wynosiła w trzecim kwartale 2020 r. ok. 5% wszystkich aktywnych użytkowników w miesiącu (ang. *monthly active users*, MAU)<sup>26</sup>. Z kolei Twitter ocenił, że w czwartym kwartale 2019 r. liczba takich kont na jego platformie stanowiła mniej niż 5%<sup>27</sup>.

## Kanały publikowania i rozpowszechniania fake newsów w internecie

Najważniejsze kanały wykorzystywane do publikowania i rozpowszechniania w sieci nieprawdziwych wiadomości to różnego rodzaju strony internetowe oraz media społecz-

<sup>24</sup> Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia, *Who Is Tweeting on Twitter: Human, Bot, or Cyborg?*, w: *Proceedings of the 26th Annual Computer Security Applications Conference*, Association for Computing Machinery, New York 2010, s. 21, <https://doi.org/10.1145/1920261.1920265>.

<sup>25</sup> E. Ferrara, *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, „First Monday” 2017, vol. 22(8), <https://doi.org/10.5210/fm.v22i8.8005>.

<sup>26</sup> *Fake Accounts*, <https://transparency.facebook.com/community-standards-enforcement#fake-accounts> [dostęp: 29 listopada 2020 r.].

<sup>27</sup> *Fiscal Year 2019 Annual Report*, s. 24, <https://investor.twitterinc.com/financial-information/annual-reports/default.aspx> [dostęp: 13 września 2020 r.].

nościowe, w szczególności serwisy społecznościowe i komunikatory internetowe. Nieprawdziwa wiadomość, która została opublikowana na jednym z nich, może być łatwo rozpowszechniana w pozostałych. Istotną rolę odgrywają tutaj również wyszukiwarki internetowe (dalej: wyszukiwarki), natomiast bardzo niewielką – poczta elektroniczna i listy mailingowe.

Fake newsy mogą być publikowane w tworzonych specjalnie w tym celu fejkowych witrynach. Ich twórcy zarabiają na umieszczonych na takich stronach reklamach<sup>28</sup>. W celu zwiększenia wiarygodności przypominają one zwykle typowe strony informacyjne lub proste portale, a ich adresy są często upodabniane do adresów URL znanych lub budzących zaufanie stron (np. mogą różnić się jedną literą w nazwie lub rozszerzeniem). Publikowane tam artykuły mają często sensacyjne lub wywołujące emocje tytuły, co ma skłonić zaciekawionego lub zaintrygowanego czytelnika do kliknięcia i zapoznania się z ich pełną treścią (ang. *clickbait*). Nieprawdziwe treści mogą stanowić tylko część zawartości takich fejkowych witryn, tj. mogą być wymieszane z informacjami prawdziwymi.

Opublikowane na stronach internetowych fake newsy bardzo łatwo jest rozpowszechnić w serwisach społecznościowych (o czym szerzej w dalszej części artykułu), które jednocześnie są wykorzystywane do generowania ruchu internetowego na tych stronach. Wklejenie linka (adresu URL) do takiej wiadomości pozwala utworzyć nowy wpis (post) z oryginalnym tytułem i leadem (wstępem do artykułu wyświetlanego pod tytułem) lub ich fragmentami oraz grafiką zaciągniętymi z oryginalnej wiadomości. Linki są również wklejane do dodawanych pod postami komentarzy, które są bardzo ważnym narzędziem wywierania wpływu w internecie, gdyż wielu użytkowników jest ciekawych opinii innych na temat danej treści i na ich podstawie może formułować własne. Komentarze stanowią także istotną informację zwrotną dla autorów opatrzonych nimi treści i mogą mieć znaczenie dla ich dalszej aktywności.

Najmniej zbadanym kanałem rozpowszechniania fake newsów są komunikatory internetowe. Służą one bowiem do przesyłania wiadomości między dwoma osobami lub między członkami licznie tworzonych tam grup. Są więc widoczne tylko dla ich nadawców i odbiorców<sup>29</sup>. W mediach społecznościowych i na stronach internetowych nieprawdziwe treści z komunikatorów są publikowane jako zrzuty ekranów zawierające wypowiedzi lub fragmenty rozmów.

Duże znaczenie dla rozpowszechniania fake newsów mają także wyszukiwarki (zasadniczo nie można opublikować ich tam bezpośrednio). Algorytmy wyszukiwarek decydują bowiem o kolejności wyświetlania w wynikach wyszukiwań linków do stron internetowych, grafiki i nagrań wideo oraz treści w niektórych serwisach społecznościowych. Prawdo-

podobieństwo kliknięcia w link przez osoby poszukujące informacji jest tym większe, im wyżej znajduje się on na liście wyników<sup>30</sup>. Nieprawdziwe treści mogą również pojawiać się w witrynach, których linki są wyświetlane przez wyszukiwarki jako linki sponsorowane (reklamy).

## Nieprawdziwe treści w serwisach społecznościowych

Interesująca zawartość, którą można personalizować, łatwość nawiązywania i podtrzymywania kontaktów, tworzenie własnej treści oraz możliwość promowania własnej aktywności – to wszystko sprawia, że dla wielu osób serwisy społecznościowe stały się ulubioną rozrywką i podstawowym źródłem pozyskiwania nowych informacji. Jednak w odróżnieniu od mediów tradycyjnych, których pracownicy pełnią funkcję „filtra” dla rozpowszechnianych poprzez nie wiadomości, w przypadku mediów społecznościowych praktycznie nie ma barier wejścia. Każdy ich użytkownik może być autorem dowolnej treści, wymaga to jedynie zarejestrowania się (utworzenia konta użytkownika). Taką treść może być bardzo szybko rozpowszechniona poprzez: udostępnianie jej przez użytkowników na swoich profilach i w grupach dyskusyjnych, odwołania i linkowanie jej w komentarzach, a także wykorzystywanie jej do tworzenia nowej treści.

Powody rozpowszechniania fake newsów w serwisach społecznościowych są następujące:

- wywoływanie u ich odbiorców emocji (zaskoczenia, gniewu, złości, irytacji, strachu czy radości), którym towarzyszy impuls do natychmiastowego podzielenia się tymi informacjami z otoczeniem;
- uznanie takich treści przez ich odbiorców za prawdziwe (i interesujące dla innych osób); prawdopodobieństwo tego zwiększają m.in.: ich zgodność z poglądami odbiorców (tzw. efekt potwierdzenia; ang. *confirmation bias*), zapoznanie się z nimi kilka lub więcej razy (tzw. efekt iluzorycznej prawdy; ang. *illusory truth effect*)<sup>31</sup> oraz to, że ich autorami są osoby znane odbiorcy, otrzymał je od takich osób lub pochodzą ze źródeł, do których ma zaufanie;
- niedokładne zapoznanie się z taką treścią, spowodowane m.in. przeciążeniem otrzymywanymi informacjami (ang. *information overload*), pośpiechem lub utrudnionymi warunkami (złe oświetlenie, małe urządzenia mobilne);
- brak umiejętności zweryfikowania prawdziwości informacji, brak chęci do tego czy też trudności związane z taką weryfikacją.

Treści nieprawdziwe są często bardziej angażujące od pozostałych informacji, tj. mogą uzyskiwać więcej polubień, komentarzy i, co szczególnie istotne, udostępnień (uznawanych za najwyższą formę zainteresowania)<sup>32</sup>. Mogą rów-

<sup>28</sup> Najbardziej znanym przykładem była działalność grupy młodych ludzi z miasta Weles w Macedonii Północnej, którzy w czasie kampanii przed wyborami prezydenckimi w Stanach Zjednoczonych w 2016 r. stworzyli co najmniej 140 witryn internetowych, na których publikowali treści wspierające Donalda Trumpa. Zob. C. Silverman, L. Alexander, *How Teens in the Balkans Are Duping Trump Supporters With Fake News*, <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo> [dostęp: 23 grudnia 2020 r.].

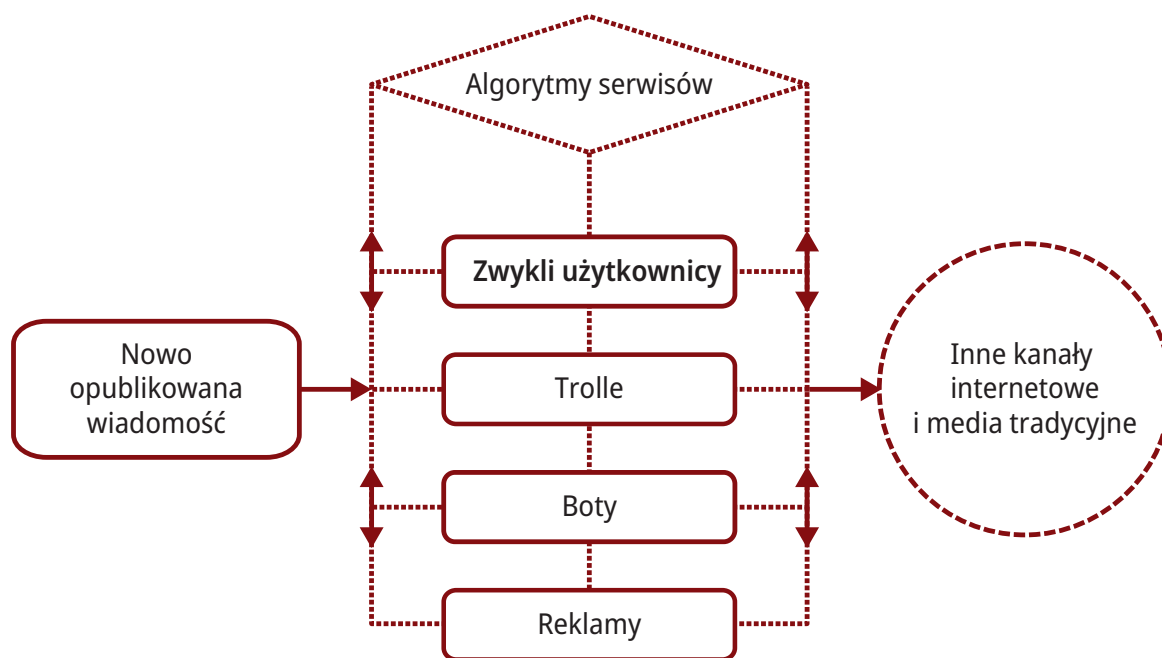
<sup>29</sup> W przypadku popularnego komunikatora WhatsApp są one także szyfrowane (co wyklucza ich moderowanie).

<sup>30</sup> Istnieje szereg sposobów wpływania na algorytmy wyszukiwarek, co szczegółowo określają wskazówki dotyczące optymalizacji dla wyszukiwarek internetowych (ang. *search engine optimization*, SEO).

<sup>31</sup> G. Pennycook, T.D. Cannon, D.G. Rand, *Prior Exposure Increases Perceived Accuracy of Fake News*, „Journal of Experimental Psychology: General” 2018, vol. 147(12), <https://doi.org/10.1037/xge0000465>.

<sup>32</sup> N. Marchal, B. Kollanyi, L.-M. Neudert, P.N. Howard, *Junk News During the EU Parliamentary Elections: Lessons From a Seven-Language*

Rysunek 1. Rozpowszechnianie fake newsów w serwisach społecznościowych



Źródło: opracowanie własne.

niez rozpowszechniać się szybciej i mieć większe zasięgi niż wiadomości prawdziwe<sup>33</sup>. Błyskawiczne uzyskiwanie dużych zasięgów w internecie jest nazywane wiralnością (ang. *viral*; wirusowy), a takie popularne treści – wiralami.

Jeżeli fake newsy są częścią prowadzonych akcji dezinformacyjnych lub wzmacniających określone narracje, to najczęściej są rozpowszechniane przez skoordynowane działania uczestniczących w nich użytkowników (głównie trolli), ich pacynek oraz botów. W terminologii spółki Facebook takie działania wprowadzające w błąd innych użytkowników co do prawdziwej tożsamości uczestniczących w nich osób i ich celów są nazywane „skoordynowanymi nieautentycznymi zachowaniami” (ang. *coordinated inauthentic behavior*, CIB) i są zabronione w jej serwisach – Facebook i Instagram<sup>34</sup>.

Dla uzyskania dużych zasięgów decydujące znaczenie mają jednak procesy organiczne, czyli wielkość zainteresowania i działania zwykłych użytkowników (rysunek 1). Większość nieprawdziwych treści nie ma bowiem dużego odbio-

ru i rozpowszechnia je niewiele osób<sup>35</sup>. Stąd ogromną rolę w tych procesach odgrywają popularni użytkownicy, czyli tzw. influencerzy (ang. *influencers*). Są to najczęściej politycy, aktorzy, muzycy, sportowcy, dziennikarze, blogerzy i osobowości medialne. Mają oni bardzo dużą, często wielomilionową, liczbę fanów czy kont obserwujących i z tego właśnie powodu trolle i operatorzy botów starają się zwrócić ich uwagę na swoje treści i zachęcić do ich udostępnienia. Robią to najczęściej przez dodawanie komentarzy pod postami influencerów i wzmiankowanie ich w swoich postach. Bardzo dużą rolę w informowaniu o fake newsach odgrywają również tworzone na ich podstawie memy.

Trolle i operatorzy botów starają się nie tylko dotrzeć z promowanymi przez siebie treściami bezpośrednio do jak największej liczby użytkowników, ale równocześnie wpływać na algorytmy serwisów. Na podstawie gromadzonych danych dotyczących zainteresowania oglądanymi treściami (polubienia, dodawane komentarze, udostępnienia) oraz kliknięć w reklamy (treści sponsorowane, promowane) algorytmy decydują o tym, czy i które pojawiające się w serwisie treści zostaną wyświetlone użytkownikom na ich spersonalizowanych stronach aktualności (nazywanych różnie przez poszczególne serwisy)<sup>36</sup>, a także jak często i w jakiej kolejno-

*Study of Twitter and Facebook*, s. 4, <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Data-Memo.pdf> [dostęp: 9 grudnia 2020 r.].

<sup>33</sup> Zob. S. Vosoughi, D. Roy, S. Aral, *The Spread of True and False News Online*, „Science” 2018, vol. 359(6380), <https://doi.org/10.1126/science.aap9559>. Badanie dotyczyło serwisu Twitter.

<sup>34</sup> W czasie akcji dezinformacyjnych lub promujących określone narracje są wykorzystywani różni aktorzy i różne narzędzia jednocześnie. Facebook podkreśla, że w przypadku wykrycia CIB o zablokowaniu powiązanych kont i usunięciu stron oraz grup decydują zachowania, a nie wkład. Spółka regularnie publikuje także raporty na ten temat. Zob. np. *October 2020 Coordinated Inauthentic Behavior Report*, <https://about.fb.com/news/2020/11/october-2020-cib-report> [dostęp: 13 listopada 2020 r.].

<sup>35</sup> A. Guess, J. Nagler, J. Tucker, *Less Than You Think: Prevalence and Predictors of Fake News Dissemination of Facebook*, „Science Advances” 2019, vol. 5(1), <https://doi.org/10.1126/sciadv.aau4586>; N. Grinberg i in., *Fake News on Twitter During the 2016 U.S. Presidential Election*, „Science” 2019, vol. 363(6425), <https://doi.org/10.1126/science.aau2706>.

<sup>36</sup> Np. News Feed w serwisie Facebook jest strumieniem treści będących mieszanką uznanych przez algorytmy serwisu za potencjalnie najbardziej interesujące nowych postów znajomych użytkownika (po zmianach wprowadzonych w 2018 r. ta kategoria treści ma

ści. Dzięki temu wydłuża się czas korzystania przez nich z serwisów, co z kolei pozwala na pokazanie im większej liczby reklam. Ruch wygenerowany przez trolle i boty wokół promowanej przez nich treści może sprawić, że algorytmy serwisów zaczną ją wyświetlać większej liczbie użytkowników, co do których zgromadzone przez nie dane wskazują, że może być ona dla nich interesująca. Algorytmy nie potrafią jednak weryfikować ich prawdziwości i pokazują w aktualnościach także zyskujące popularność treści niskiej jakości<sup>37</sup>.

Do rozpowszechniania fake newsów mogą też zostać wykorzystane reklamy, jeżeli nie zakazują tego regulaminy serwisów. Serwisy społecznościowe dysponują prostymi narzędziami do zarządzania reklamami, a ich publikowanie jest stosunkowo tanie. Oferują one także możliwość ich wyświetlania wąsko definiowanym grupom użytkowników, czyli tzw. targetowanie/mikrotargetowanie (ang. *targeted advertising*)<sup>38</sup>.

Z serwisów społecznościowych fake newsy mogą szybko trafiać do innych kanałów internetowych, a także mediów tradycyjnych. Są one bowiem ważnym źródłem pozyskiwania interesujących informacji i tematów przez dziennikarzy<sup>39</sup>.

Oprócz możliwości personalizowania treści i działań algorytmów rozpowszechnianiu nieprawdziwych treści w serwisach sprzyja także homofilia, czyli skłonność do nawiązywania i utrzymywania relacji z osobami o podobnych cechach, (znacznie wyższy priorytet niż dawniej), treści opublikowanych na polubionych przez niego stronach, w grupach dyskusyjnych, których jest członkiem, a także reklam.

<sup>37</sup> Rozpowszechnianie w serwisach społecznościowych wprowadzających w błąd treści z wykorzystaniem algorytmów tych serwisów i automatyzacji jest nazywane propagandą obliczeniową (ang. *computational propaganda*). To określenie zostało spopularyzowane przez interdyscyplinarną grupę naukowców z Oxford Internet Institute na Uniwersytecie Oksfordzkim, realizujących w wybranych państwach (w tym w Polsce) projekt badawczy pod nazwą „Computational Propaganda Project”. Na jego stronie internetowej (<https://comprop.oii.ox.ac.uk>) znajduje się wiele wartościowych opracowań na temat dezinformacji w internecie.

<sup>38</sup> Np. Facebook oferuje swoim klientom możliwość definiowania grup docelowych odbiorców wyświetlanych w serwisie reklam za pomocą tysięcy kombinacji określających ich miejsce zamieszkania, dane demograficzne, zainteresowania i zachowania. Zob. <https://www.facebook.com/business/ads/ad-targeting> [dostęp: 1 grudnia 2020 r.]. Jeszcze większe możliwości daje tzw. profilowanie psychograficzne, pozwalające na precyzyjne dotarcie z odpowiednio przygotowaną treścią do osób, które będą na zawarty w niej przekaz najbardziej podatne. Takie usługi świadczyła brytyjska spółka Cambridge Analytica na podstawie danych o ok. 87 mln użytkowników Facebooka i ich znajomych, które zebrano za pomocą specjalnej aplikacji. Wywołało to w 2018 r. jeden z największych skandali w historii mediów społecznościowych. Zob. N. Badshah, *Facebook to Contact 87 Million Users Affected by Data Breach*, 8 April 2018, <https://www.theguardian.com/technology/2018/apr/08/facebook-to-contact-the-87-million-users-affected-by-data-breach> [dostęp: 23 grudnia 2020 r.].

<sup>39</sup> Coraz częściej zwraca się uwagę na rolę, jaką w rozpowszechnianiu fake newsów odgrywają dziennikarze i media tradycyjne. Zob. np. Y. Tsifti i in., *Causes and Consequences of Mainstream Media Dissemination of Fake News: Literature Review and Synthesis*, „Annals of the International Communication Association” 2020, vol. 44(2), <https://doi.org/10.1080/23808985.2020.1759443>. Chociaż działają często z dobrymi intencjami skrytykowania lub zdementowania nieprawdziwych treści, to jednocześnie nagłaśniają je, informując o nich osoby, które o nich nigdy by się nie dowiedziały.

wartościach, zainteresowaniach czy poglądach oraz bycia członkiem grup skupiających takie osoby. Jednym z następstw homofilii jest zjawisko tzw. bańki filtrującej<sup>40</sup> (ang. *filter bubble*; zwyczajowo: bańka informacyjna). Polega ono na dopływie informacji zgodnych z posiadanymi poglądami. Wzmacniają je także algorytmy nadające wysoki priorytet wyświetlania w aktualnościach treściom, których autorami są znajomi lub które wzbudziły zainteresowanie znajomych użytkownika<sup>41</sup>. Tworzeniu się bańek sprzyja także usuwanie z sieci kontaktów znajomych, którzy ujawnili odmienne poglądy, czy też ignorowanie lub ukrywanie w aktualnościach ich wkładu. Kolejnym ważnym zjawiskiem są tzw. komory pogłosowe (ang. *echo chambers*), czyli środowiska, w których komunikują się ze sobą osoby o podobnych poglądach. W serwisach społecznościowych występuje ono najczęściej w grupach dyskusyjnych (zwłaszcza zamkniętych). Komory pogłosowe sprzyjają szybkiemu rozpowszechnianiu się tych fake newsów, które są zgodne z poglądami znajdujących się w tych środowiskach osób, oraz zwiększają ich przekonanie o prawdziwości takich treści<sup>42</sup>.

### Ograniczanie liczby i zasięgów fake newsów w serwisach społecznościowych

Amerykańskie spółki, które są właścicielami większości najpopularniejszych serwisów społecznościowych na świecie, były początkowo ostrożne co do bardziej zdecydowanego moderowania treści na swoich platformach. Po pierwsze, obawiały się zarzutów o stosowanie cenzury i ograniczanie wolności słowa. Zamieszczanie nieprawdziwych treści co do zasady nie jest bowiem niezgodne z regulaminami serwisów, a tym samym nie jest to dostateczny warunek usuwania takiej treści. Rezerwę wobec podejmowania działań w tym obszarze najpełniej oddaje często powtarzane stwierdzenie członków najwyższego kierownictwa Facebooka i Twittera, że nie chcą być „arbitrami prawdy” (ang. *arbiters of the truth*)<sup>43</sup>. Zapisy 230. sekcji amerykańskiej ustawy Communications Decency Act z 1996 r. chronią te spółki (z kilkoma wyjątkami) przed odpowiedzialnością prawną za treści publikowane przez użytkowników na ich platformach. Jednocześnie są to przedsiębiorstwa prywatne, a więc nie dotyczy ich pierwsza poprawka do Konstytucji Stanów Zjednoczonych (zakazuje ona bowiem ograniczania wolności słowa przez rząd federalny, władze stanowe i lokalne)<sup>44</sup>. Mogą więc same podej-

<sup>40</sup> K. Sharma i in., *Combating Fake News: A Survey on Identification and Mitigation Techniques*, „ACM Transactions on Intelligent Systems and Technology” 2019, vol. 10(3), s. 5, <https://doi.org/10.1145/3305260>.

<sup>41</sup> E. Bakshy, S. Messing, L.A. Adamic, *Exposure to Ideologically Diverse News and Opinion on Facebook*, „Science” 2015, vol. 348(6239), <https://doi.org/10.1126/science.aaa1160>.

<sup>42</sup> K. Shu i in., *Fake News Detection on Social Media: A Data Mining Perspective*, „ACM SIGKDD Explorations Newsletter” 2017, vol. 19(1), s. 25, <https://doi.org/10.1145/3137597.3137600>.

<sup>43</sup> P.M. Barrett, *Tackling Domestic Disinformation: What the Social Media Companies Need to Do*, NYU Stern Center for Business and Human Rights, New York 2019, s. 17.

<sup>44</sup> D.L. Hudson, *In the Age of Social Media Expand Reach of the First Amendment*, American Bar Association, [https://www.americanbar.org/groups/crsj/publications/human\\_rights\\_magazine\\_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-social-media-first-amendment](https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-social-media-first-amendment) [dostęp: 30 grudnia 2020 r.].



mować decyzje, jakie treści akceptują w swoich serwisach, a jakich nie<sup>45</sup>.

Po drugie, zmniejszanie na platformach liczby i widoczności fake newsów jest niekorzystne dla opartych na reklamach modeli biznesowych ich właścicieli. Wymagają one bowiem maksymalizowania liczby odwiedzin i czasu spędzanego w serwisach przez jak największą liczbę użytkowników. To z kolei pozwala zyskownie sprzedawać ich uwagę reklamodawcom. Tymczasem niskiej jakości, sensacyjne treści są atrakcyjne dla wielu osób. Do tego boty społecznościowe, cyborgi, trolle i ich pacyнки zwiększają liczbę aktywnych użytkowników serwisów, która dla akcjonariuszy i inwestorów jest ważnym miernikiem oceny realizacji celów dotyczących ich wzrostu.

Po trzecie, moderowanie dodawanych przez użytkowników treści zwiększa koszty działalności. Pomimo coraz szerszego wykorzystywania sztucznej inteligencji (zwłaszcza uczenia maszynowego) procesów blokowania kont użytkowników i treści naruszających regulaminy serwisów nie można całkowicie zautomatyzować. Zwłaszcza treści dodawane przez trolle oraz niektóre formaty treści (np. nagrania wideo) wymagają ich sprawdzania przez ludzi. Generuje to koszty zatrudniania moderatorów lub outsourcingu takich usług<sup>46</sup>.

Nacisk opinii publicznej, wszczęte dochodzenia parlamentarne, nowe regulacje dyskutowane lub wprowadzane w niektórych państwach, niezadowolone części użytkowników i reklamodawców, a także straty wizerunkowe związane z łatwym wykorzystywaniem mediów społecznościowych w procesach dezinformacji sprawiły, że w latach 2016–2020 właściciele wielu z nich podjęli działania mające utrudnić aktywność na ich platformach autorom i propagatorom fake newsów oraz ograniczyć liczbę i zasięg takich treści. Podstawą do tego stały się ich stale aktualizowane, uszczegóławiane i surowiej egzekwowane polityki, zasady i regulaminy<sup>47</sup>. Do najważniejszych takich działań należały:

- utrudnianie tworzenia fałszywych kont<sup>48</sup>;
- blokowanie fałszywych kont i usuwanie ich wkładu;
- czasowe blokowanie lub usuwanie kont użytkowników naruszających regulaminy serwisów;

<sup>45</sup> D. Wakabayashi, *Legal Shield for Social Media is Targeted by Lawmakers*, <https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html> [dostęp: 7 grudnia 2020 r.].

<sup>46</sup> Mogą tutaj również wystąpić inne dodatkowe koszty i negatywne skutki uboczne, np. ugoda o wartości 52 mln dolarów zawarta przez Facebooka w maju 2020 r. zwróciła uwagę na fakt, że weryfikowane treści, zawierające m.in. mowę nienawiści, narażają moderatorów na ryzyko wystąpienia zespołu stresu pourazowego. Zob. *Facebook to Pay \$52m to Content Moderators Over PTSD*, <https://www.bbc.com/news/technology-52642633> [dostęp: 5 grudnia 2020 r.].

<sup>47</sup> Np. *Rules and Policies* Twittera (<https://help.twitter.com/en/rules-and-policies>), *Regulamin i zasady obowiązujące na Facebooku* (<https://www.facebook.com/policies>) czy *Wytyczne dla społeczności* serwisu YouTube (<https://www.youtube.com/howyoutubeworks/policies/community-guidelines>).

<sup>48</sup> W II kwartale 2020 r. liczba udaremnionych prób utworzenia fałszywych kont w serwisie Facebook wyniosła 1,5 mld, podczas gdy w tym samym kwartale średnie MAU wynosiło 2,7 mld. W zdecydowanej większości były to jednak zautomatyzowane próby masowego tworzenia fałszywych kont w celu rozpowszechniania spamu, a nie fake newsów. Zob. <https://transparency.facebook.com/community-standards-enforcement#fake-accounts> i <https://investor.fb.com/investor-events/event-details/2020/Facebook-Q2-2020-Earnings-default.aspx> [dostęp: 16 września 2020 r.].

- usuwanie lub ukrywanie treści naruszających regulaminy oraz utrudnianie lub uniemożliwianie powtórnego publikowania takich samych lub zmodyfikowanych wersji takich treści<sup>49</sup>;
- zatrudnienie większej liczby moderatorów ze znajomością większej liczby języków obcych<sup>50</sup>;
- ułatwianie użytkownikom zgłaszania kont i treści, które w ich ocenie mogą naruszać regulaminy;
- nawiązanie współpracy z partnerami zewnętrznymi sprawdzającymi odpłatnie prawdziwość wskazywanych przez serwisy treści (wciąż głównie manualnie<sup>51</sup>, chociaż również tutaj zaczyna być wykorzystywana sztuczna inteligencja<sup>52</sup>);
- oznaczanie treści, których prawdziwość została zakwestionowana przez partnerów zewnętrznych lub moderatorów, poprzez umieszczanie przy nich ostrzeżeń i dodatkowych informacji pochodzących ze źródeł o uznanej wiarygodności<sup>53</sup> (zapewnia to szerszy kontekst i zmniejsza skłonność użytkowników do ich udostępniania)<sup>54</sup>;
- wyświetlanie treści, których prawdziwość została zakwestionowana przez zewnętrznych parterów, mniej

<sup>49</sup> Np. serwis YouTube w II kwartale 2020 r. usunął 11,4 mln nagrań wideo i 2,1 mld komentarzy, Twitter w II połowie 2019 r. – 2,9 mln tweetów, a TikTok w I połowie 2020 r. – 105 mln nagrań wideo. W przypadku wszystkich trzech serwisów były to wielkości znacznie większe niż w poprzednich okresach. Zob. *The Great Clean-Up*, „The Economist”, 24–30 October 2020, s. 18. Każda spółka ma swoje zasady i wytyczne dotyczące moderowania treści, co najlepiej oddaje fakt, że taki sam post opublikowany w różnych serwisach może się spotkać z różnymi reakcjami ich moderatorów. Zob. np. *US Election: Twitter Hides Trump Tweet About 'Disappearing' Lead*, <https://www.bbc.com/news/technology-54809165> [dostęp: 1 grudnia 2020 r.].

<sup>50</sup> W 2018 r. liczba osób moderujących treści w serwisie Facebook wyniosła ok. 15 tys. Byli to pracownicy zarówno spółki, jak i podmiotów zewnętrznych działających na jej zlecenie. Moderatorzy znali ponad 50 języków obcych i pracowali w ponad 20 lokalizacjach w każdej strefie czasowej. Zob. *Hard Questions: Who Reviews Objectionable Content on Facebook – And Is the Company Doing Enough to Support Them?*, <https://about.fb.com/news/2018/07/hard-questions-content-reviewers> [dostęp: 24 września 2020 r.]. W 2020 r. liczba moderatorów pracujących dla Facebooka wyniosła już ok. 35 tys. Zob. *The Great...*, *op. cit.*, s. 19.

<sup>51</sup> Istotnym ograniczeniem tego rozwiązania jest czas reakcji. Sprawdzenie prawdziwości informacji przez zewnętrznych partnerów trwa średnio od kilku do kilkunastu godzin, w którym to czasie potencjalnie nieprawdziwa treść dociera do kolejnych użytkowników.

<sup>52</sup> *Why AI Live-Checked the 2020 Presidential Elections*, <https://www.bbc.com/news/av/technology-54658206> [dostęp: 10 grudnia 2020 r.].

<sup>53</sup> Szczególnie duży rozgłos zyskało rozpoczęte w maju 2020 r. oznaczanie w ten sposób przez serwis Twitter tweetów prezydenta Stanów Zjednoczonych Donalda Trumpa, <https://www.cnbc.com/2020/05/26/twitter-fact-checks-trump-slaps-warning-labels-on-his-tweets-about-mail-in-ballots.html> [dostęp: 28 grudnia 2020 r.].

<sup>54</sup> Prawdopodobieństwo udostępnienia dalej treści zwiększają m.in. przekonanie o ich prawdziwości oraz ich zgodność z posiadanymi przekonaniem. Zob. T. Buchanan, *Why Do People Spread False Information Online? The Effects of Message and Viewer Characteristics on Self-Reported Likelihood of Sharing Social Media Disinformation*, „Plos ONE” 2020, vol. 15(10), <https://doi.org/10.1371/journal.pone.0239666>.

szej liczbie użytkowników w ich aktualnościach<sup>55</sup> oraz ograniczanie możliwości ich dalszego udostępniania;

- zakazywanie (tymczasowe lub stałe) publikowania niektórych rodzajów reklam (najczęściej dotyczących zdrowia i polityki), zwiększenie kontroli nad ich treścią, a także dostępności informacji o niektórych reklamodawcach<sup>56</sup>;
- eliminowanie zachęt finansowych do wymyślania i rozpowszechniania fake newsów poprzez nieuwzględnianie witryn internetowych zawierających takie treści w systemach sprzedaży reklam;
- inicjatywy podejmowane w związku z oczekiwanym nasileniem zjawiska rozpowszechniania nieprawdziwych treści podczas wyborów, takie jak uruchomienie przez Facebooka w latach 2018 i 2019 dedykowanych zespołów pracowników i całodobowych centrów operacji wyborczych w biurach spółki w Menlo Park, Dublinie i Singapurze<sup>57</sup>;
- zakaz publikowania i rozpowszechniania filmów stanowiących *deepfakes*;
- inicjatywy zwiększające wiedzę na temat zagrożeń związanych z nieprawdziwymi treściami i sposobów ich rozpoznawania (kampanie informacyjne, szkolenia dla dziennikarzy).

W 2018 r. Facebook, Twitter, Google (właściciel serwisu YouTube), Mozilla i niektóre stowarzyszenia branżowe z sektora reklamy podpisały Unijny kodeks postępowania w zakresie zwalczania dezinformacji (ang. EU Code of Practice on Disinformation)<sup>58</sup>. W 2019 r. do sygnatariuszy kodeksu dołączył Microsoft, a na początku 2020 r. TikTok<sup>59</sup>. Kodeks zawiera 21 zobowiązań dotyczących m.in. zwiększenia kontroli nad reklamami (zwłaszcza politycznymi) i ograniczenia niewłaściwego wykorzystywania botów. Jednak jego sygnatariusze realizują jedynie te zobowiązania, które dotyczą ich produktów i usług, wynikają z ich pozycji w łańcuchu wartości dostarczania treści oraz uwzględniają ich możliwości techniczne w państwach należących do Europejskiego Obszaru Gospodarczego<sup>60</sup>.

<sup>55</sup> W spółce Facebook to podejście nosi nazwę *Remove, Reduce, Inform* (usuwamy, ograniczamy, redukujemy), zob. szerzej artykuł Jakuba Turowskiego: J. Turowski, *Walka, która nigdy się nie skończy – jak Facebook ogranicza rozprzestrzenianie się dezinformacji*, w: *Zjawisko dezinformacji w dobie rewolucji cyfrowej. Państwo. Społeczność. Polityka. Biznes*, NASK Państwowy Instytut Badawczy, Warszawa 2019, s. 61–64. W swoich sprawozdaniach dla Komisji Europejskiej spółka podała, że takie działania spowodowały ograniczenie wyświetleń zakwestionowanych treści przeciętnie o ponad 80%. Zob. *Facebook Baseline Report on Implementation of the Code of Practice on Disinformation*, s. 2, [https://ec.europa.eu/information\\_society/newsroom/image/document/2019-5/facebook\\_baseline\\_report\\_on\\_implementation\\_of\\_the\\_code\\_of\\_practice\\_on\\_disinformation\\_CF161D11-9A54-3E27-65D58168CAC40050\\_56991.pdf](https://ec.europa.eu/information_society/newsroom/image/document/2019-5/facebook_baseline_report_on_implementation_of_the_code_of_practice_on_disinformation_CF161D11-9A54-3E27-65D58168CAC40050_56991.pdf) [dostęp: 10 września 2020 r.].

<sup>56</sup> Np. poprzez uruchomienie przez Facebooka biblioteki reklam: <https://www.facebook.com/ads/library>.

<sup>57</sup> W anglojęzycznej terminologii IT takie rozwiązanie organizacyjne jest określane jako *war room* (pokój wojenny).

<sup>58</sup> Dokument jest dostępny także w języku polskim na stronie: <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation> [dostęp: 2 września 2020 r.].

<sup>59</sup> <https://ec.europa.eu/digital-single-market/en/news/roadmaps-implementation-code-practice-disinformation> [dostęp: 5 sierpnia 2020 r.].

<sup>60</sup> Dużą wartość informacyjną o działaniach podjętych przez spółki w celu ograniczenia dezinformacji mają dokumenty opubliko-

Ocena skuteczności poszczególnych interwencji mających zmniejszyć liczbę nieprawdziwych treści jest trudna i wymaga zastosowania metod eksperymentalnych<sup>61</sup>. Wydaje się, że mogły być one przynajmniej częściowo skuteczne, chociaż występują tutaj duże różnice między serwisami<sup>62</sup> (np. na Facebooku po 2016 r. liczba interakcji użytkowników z nieprawdziwymi treściami znacząco spadła<sup>63</sup>). Pomimo niewątpliwej poprawy zakresu i skuteczności tych działań są przedmiotem nieustających dyskusji i częstej krytyki. Oparte na sztucznej inteligencji systemy serwisów społecznościowych mające na celu zapobieganie i wykrywanie szkodliwej działalności botów nie zawsze nadążają za ich rosnącym zaawansowaniem i taktyką ich operatorów. Chociaż zwiększenie liczby moderatorów przez największe serwisy pomaga weryfikować i usuwać więcej kont i ich wkładu w większej liczbie języków, to nadal często przegrywają oni z pomysłowością, zawziętością i pracowitością płatnych trolli oraz trollujących użytkowników. Problemami są także częsty brak reakcji moderatorów lub długi czas oczekiwania na nie, a także ich błędne decyzje, zgodnie z którymi działalność zgłaszanych kont nie narusza regulaminów<sup>64</sup>.

Z kolei kwestia działań, które mogą być podejmowane przez władze publiczne w celu ograniczania zjawiska fake newsów, wymaga omówienia już w oddzielnym artykule. Regulacje prawne odnoszące się do rozpowszechniania nieprawdziwych treści w internecie wprowadziły do grudnia 2020 r. m.in.: Francja, Niemcy, Indonezja, Kenia, Malediwy, Rosja, Singapur, Tajlandia i Uganda<sup>65</sup>. Bardzo duże znaczenie dla zwiększania odporności na fake newsy mają również krajowe systemy oświatowe, w tym same programy nauczania (w szczególności edukacja medialna).

## Podsumowanie

Rozpowszechnianie nieprawdziwych treści w mediach społecznościowych jako sposób realizacji swoich celów przez osoby, organizacje oraz państwa okazało się zjawiskiem trwałym. Należy założyć, że fake newsy w dalszym ciągu będą

wane na stronach Komisji Europejskiej: załącznik do porozumienia z opisami i linkami do ich najlepszych praktyk, ich okresowe sprawozdania oraz raport z oceny wdrażania kodeksu: <https://ec.europa.eu/digital-single-market/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019> [dostęp: 10 sierpnia 2020 r.]; <https://ec.europa.eu/digital-single-market/en/news/study-assessment-implementation-code-practice-disinformation> [dostęp: 10 sierpnia 2020 r.].

<sup>61</sup> G. Pennycook, D.G. Rand, *Evaluating Interventions to Fight Misinformation*, <https://misinforeview.hks.harvard.edu/article/tackling-misinformation-what-researchers-could-do-with-social-media-data> [dostęp: 30 grudnia 2020 r.].

<sup>62</sup> S. Bay, A. Dek, I. Dek, R. Fredheim, *op. cit.*, s. 27–32.

<sup>63</sup> H. Allcott, M. Gentzkow, C. Yu, *Trends in the Diffusion of Misinformation on Social Media*, „Research and Politics” 2019, vol. 6(2), <https://doi.org/10.1177/2053168019848554>.

<sup>64</sup> S. Bay, A. Dek, I. Dek, R. Fredheim, *op. cit.*, s. 4.

<sup>65</sup> W niektórych państwach problemy z fake newsami i dezinformacją w sieci zostały jednak wykorzystane przez władze jako pretekst do ograniczenia wolności słowa i zwiększenia cenzury w internecie. Zob. A. Shabnaz, *The Rise of Digital Authoritarianism*, Freedom House, <https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism> [dostęp: 23 grudnia 2020 r.].

zanieczyszczają serwisy społecznościowe. Takie treści (wciąż w niedostatecznym stopniu zbadane) są bowiem atrakcyjne dla wielu osób i istnieje dobrze funkcjonujący rynek usług manipulacyjnych w internecie. Zdaniem wielu ekspertów w tej sytuacji kluczowego znaczenia nabiera ostatnia linia obrony: kształtowanie świadomości, postaw oraz kompetencji cyfrowych konsumentów fake newsów, których można (i trzeba) uczyć. Najczęściej wymienia się tutaj krytyczne myślenie, umiejętność zachowywania emocjonalnego dystansu oraz sceptycyzm co do czytanych i oglądanych treści. Ważny jest także nawyk podejmowania dodatkowego wysiłku weryfikowania budzących wątpliwości treści (czasem z wykorzystaniem kilku mniej znanych funkcji wyszukiwarek, takich jak odwrotne wyszukiwanie grafiki). W sytuacji, kiedy ekspozycja na fake newsy jest warunkiem korzystania z mediów społecznościowych, każdy powinien być trochę fact-checkerem.

## Bibliografia

- Allcott H., Gentzkow M., Yu C., *Trends in the Diffusion of Misinformation on Social Media*, „Research and Politics” 2019, vol. 6(2), <https://doi.org/10.1177/2053168019848554>.
- Bakshy E., Messing S., Adamic L.A., *Exposure to Ideologically Diverse News and Opinion on Facebook*, „Science” 2015, vol. 348(6239), <https://doi.org/10.1126/science.aaa1160>.
- Barrett P.M., *Tackling Domestic Disinformation: What the Social Media Companies Need to Do*, NYU Stern Center for Business and Human Rights, New York 2019.
- Bay S., Dek A., Dek I., Fredheim R., *Social Media Manipulation 2020*, NATO StratCOM COE, Riga 2020.
- Bradshaw S., Howard P.N., *The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation*, Computational Propaganda Research Project, University of Oxford, Oxford 2019.
- Buchanan T., *Why Do People Spread False Information Online? The Effects of Message and Viewer Characteristics on Self-Reported Likelihood of Sharing Social Media Disinformation*, „Plos ONE” 2020, vol. 15(10), <https://doi.org/10.1371/journal.pone.0239666>.
- Chan M.S., Jones C.R., Jamieson K.H., Albarracín D., *Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation*, „Psychological Science” 2017, vol. 28(11), <https://doi.org/10.1177/0956797617714579>.
- Chu Z., Gianvecchio S., Wang H., Jajodia S., *Who is Tweeting on Twitter: Human, Bot, or Cyborg?*, w: *Proceedings of the 26th Annual Computer Security Applications Conference*, Association for Computing Machinery, New York 2010, <https://doi.org/10.1145/1920261.1920265>.
- Ferrara E., *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, „First Monday” 2017, vol. 22(8), <https://doi.org/10.5210/fm.v22i8.8005>.
- Flynn D.J., Nyhan B., Reifler J., *The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics*, „Political Psychology” 2017, vol. 38, <https://doi.org/10.1111/pops.12394>.
- Grinberg N. i in., *Fake News on Twitter During the 2016 U.S. Presidential Election*, „Science” 2019, vol. 363(6425), <https://doi.org/10.1126/science.aau2706>.
- Guess A., Nagler J., Tucker J., *Less Than You Think: Prevalence and Predictors of Fake News Dissemination of Facebook*, „Science Advances” 2019, vol. 5(1), <https://doi.org/10.1126/sciadv.aau4586>.
- King G., Pan J., Roberts M.E., *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument*, „American Political Science Review” 2017, vol. 111(3), <https://doi.org/10.1017/s0003055417000144>.
- Lazer D.M.J. i in., *The Science of Fake News*, „Science” 2018, vol. 539(6380), <https://doi.org/10.1126/science.aao2998>.
- October 2020 Coordinated Inauthentic Behavior Report*, Facebook, <https://about.fb.com/wp-content/uploads/2020/11/October-2020-CIB-Report.pdf>.
- Pennycook G., Cannon T.D., Rand D.G., *Prior Exposure Increases Perceived Accuracy of Fake News*, „Journal of Experimental Psychology: General” 2018, vol. 147(12), <https://doi.org/10.1037/xge0000465>.
- Shao C. i in., *The Spread of Low-Credibility Content by Social Bots*, „Nature Communications” 2018, vol. 9(1), <https://doi.org/10.1038/s41467-018-06930-7>.
- Sharma K. i in., *Combating Fake News: A Survey on Identification and Mitigation Techniques*, „ACM Transactions on Intelligent Systems and Technology” 2019, vol. 10(3), s. 5, <https://doi.org/10.1145/3305260>.
- Shu K. i in., *Fake News Detection on Social Media: A Data Mining Perspective*, „ACM SIGKDD Explorations Newsletter” 2017, vol. 19(1), <https://doi.org/10.1145/3137597.3137600>.
- The Great Clean-Up*, „The Economist”, 24–30 October 2020.
- Tsfati Y. i in., *Causes and Consequences of Mainstream Media Dissemination of Fake News: Literature Review and Synthesis*, „Annals of the International Communication Association” 2020, vol. 44(2), <https://doi.org/10.1080/23808985.2020.1759443>.
- Tucker J.A. i in., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, 2018, <https://eprints.lse.ac.uk/87402/1/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf> [dostęp: 2 sierpnia 2020 r.].
- Turowski J., *Walka, która nigdy się nie skończy – jak Facebook ogranicza rozprzestrzenianie się dezinformacji*, w: *Zjawisko dezinformacji w dobie rewolucji cyfrowej. Państwo. Społeczeństwo. Polityka. Biznes*, NASK Państwowy Instytut Badawczy, Warszawa 2019.
- Vosoughi S., Roy D., Aral S., *The Spread of True and False News Online*, „Science” 2018, vol. 359(6380), <https://doi.org/10.1126/science.aap9559>.
- Yang K.-C. i in., *Arming the Public with Artificial Intelligence to Counter Social Bots*, „Human Behavior and Emerging Technologies” 2018, vol. 1(1), <https://doi.org/10.1002/hbe2.115>.
- Zhang J., Zhang R., Zhang Y., Yan G., *The Rise of Social Botnets: Attacks and Countermeasures*, „IEEE Transactions on Dependable and Secure Computing” 2016, vol. 16(6), <https://doi.org/10.1109/TDSC.2016.2641441>.
- Zhuravskaya E., Petrova M., Enikolopov R., *Political Effects of the Internet and Social Media*, „Annual Review of Economics” 2020, vol. 12, <https://doi.org/10.1146/annurev-economics-081919-050239>.

## Strony internetowe

- <https://about.fb.com>  
<https://cnn.com>  
<https://comprop.oii.ox.ac.uk>  
<https://ec.europa.eu>  
<https://freedomhouse.org>  
<https://investor.twitterinc.com>  
<https://misinforeview.hks.harvard.edu>  
<https://oko.press>

<https://sciencemediahub.eu>  
<https://www.americanbar.org>  
<https://www.bbc.com>  
<https://www.buzzfeednews.com>  
<https://www.cnn.com>  
<https://www.facebook.com>  
<https://www.ft.com>

<https://www.nytimes.com>  
<https://www.reuters.com>  
<https://www.snopes.com>  
<https://www.technologyreview.com>  
<https://www.who.int>  
<https://www.youtube.com>  
<http://theguardian.com>

---

**Zespół redakcyjny:**

Dobromir Dziewulak (redaktor naczelny), Adrian Grycuk (sekretarz redakcji; tel. +48 22 694 17 50,  
e-mail: [adrian.grycuk@sejm.gov.pl](mailto:adrian.grycuk@sejm.gov.pl)), Piotr Russel, Piotr Chybalski  
Biuro Analiz Sejmowych Kancelarii Sejmu, ul. Zagórna 3, 00-441 Warszawa

---

Seria Analizy BAS wydawana jest wyłącznie w wersji elektronicznej.  
Czasopismo jest dostępne w bazach danych BazEkon, C.E.E.O.L., EBSCO Legal Source oraz w serwisie informacyjnym EMIS.  
Więcej informacji na stronie stronach [www.sejm.gov.pl](http://www.sejm.gov.pl) i [www.bas.sejm.gov.pl](http://www.bas.sejm.gov.pl) w dziale publikacje.

---