Fides, Ratio et Patria. Studia Toruńskie 24/2026
ISSN 2392-1463; e-ISSN 2956-3267
s. 123–142

https://doi.org/10.56583/frp.3292

Natalia Olszanecka[*]
Uniwersytet Mikołaja Kopernika
https://orcid.org/0000-0003-3601-8280

Sztuczna inteligencja i nowa era dezinformacji: przekształcenia komunikacji politycznej w perspektywie społeczno-politycznej

Streszczenie

Rozwój generatywnej sztucznej inteligencji przesunął granice tego, co uważano za możliwe w produkcji fałszywych treści politycznych. Modele językowe pozwalają wytwarzać przekonujące teksty na skalę wcześniej nieosiągalną, syntetyzery głosu i obrazu udostępniły technologię *deepfake* szerokim grupom użytkowników, a systemy rekomendacyjne platform społecznościowych nadal decydują o zasięgu treści. Niniejszy artykuł zestawia dorobek badań politologicznych, komunikacyjnych i psychologicznych z analizą kilku studiów przypadków z lat 2023–2025: kampanii parlamentarnej na Słowacji, prawyborów w stanie New Hampshire, kampanii parlamentarnej w Polsce, wyborów prezydenckich na Tajwanie oraz w Argentynie, a także rosyjskiej operacji wpływu „Doppelganger". Rozważania prowadzą do wniosku, że najpoważniejsze konsekwencje generatywnej AI ujawniają się w długofalowej erozji zaufania do instytucji, mediów i procedur demokratycznych. W tekście omówiono aktualne mechanizmy przeciwdziałania: rozporządzenie o sztucznej inteligencji (AI Act), akt o usługach cyfrowych (DSA), inicjatywy edukacji medialnej oraz technologie znakowania i wykrywania treści syntetycznych. Wskazano także luki badawcze,

[*] **dr Natalia Olszanecka** – adiunkt na Uniwersytecie Mikołaja Kopernika w Toruniu, badaczka związana z obszarem nauk o bezpieczeństwie. W swojej pracy naukowej koncentruje się na problematyce bezpieczeństwa międzynarodowego, militaryzmu, stosunków cywilno-wojskowych oraz przemian politycznych we współczesnej Rosji. Jest autorką publikacji dotyczących m.in. rosyjskich elit władzy, prywatnych firm wojskowych, rosyjskiej dezinformacji oraz bezpieczeństwa NATO.

Licencja CC BY-NC-ND 4.0

m.in. niedostatek analiz nad komunikatorami szyfrowanymi oraz nad zjawiskiem „dywidendy kłamcy".

Słowa kluczowe: sztuczna inteligencja, dezinformacja, *deepfake*, kampania wyborcza, demokracja, media społecznościowe, polaryzacja, wojna informacyjna

ARTIFICIAL INTELLIGENCE AND THE NEW ERA OF DISINFORMATION: TRANSFORMATIONS IN POLITICAL COMMUNICATION FROM A SOCIO-POLITICAL PERSPECTIVE

Summary

The development of generative artificial intelligence has shifted the boundaries of what was considered possible in the production of false political content. Language models make it possible to generate convincing texts on a previously unattainable scale, voice and image synthesizers have made deepfake technology available to broad groups of users, and recommender systems on social media platforms continue to determine the reach of content. This article brings together research in political science, communication studies, and psychology with an analysis of several case studies from 2023–2025: the parliamentary campaign in Slovakia, the New Hampshire primary, the parliamentary campaign in Poland, the presidential elections in Taiwan and Argentina, and the Russian influence operation "Doppelganger". The discussion leads to the conclusion that the most serious consequences of generative AI emerge in the long-term erosion of trust in institutions, media, and democratic procedures. The article discusses current countermeasures: the Artificial Intelligence Act (AI Act), the Digital Services Act (DSA), media literacy initiatives, and technologies for labeling and detecting synthetic content. It also identifies research gaps, including the shortage of studies on encrypted messaging apps and on the phenomenon of the "liar's dividend".

Keywords: artificial intelligence, disinformation, deepfake, election campaign, democracy, social media, polarization, information warfare

~ • ~

Późnym wieczorem 28 września 2023 roku, dwa dni przed słowackimi wyborami parlamentarnymi, na Facebooku pojawił się plik dźwiękowy. Nagranie rzekomo rejestrowało rozmowę Michala Šimečki, lidera proeuropejskiej partii Progresívne Slovensko, z dziennikarką dziennika *Denník N*, Moniką Tódovą, na temat kupowania głosów w społeczności romskiej. Materiał wypłynął w okresie ciszy wyborczej, gdy redakcje miały ograniczone prawo do komentowania kam-

panii. Robert Fico wygrał wybory, Šimečka przegrał, a debata publiczna zawrzała: czy właśnie zobaczyliśmy „pierwsze wybory rozstrzygnięte przez *deepfake*"?[1].

Podobne pytania powtórzyły się w kolejnych miesiącach. 22 stycznia 2024 roku, w przededniu prawyborów Partii Demokratycznej w stanie New Hampshire, około 5 000 do 25 000 wyborców odebrało nagrania telefoniczne z syntetycznie wygenerowanym głosem prezydenta Joe Bidena, wzywającym do niegłosowania i „zachowania głosu na listopad". Rok później Federalna Komisja Łączności nałożyła na politycznego konsultanta Steve'a Kramera karę w wysokości 6 milionów dolarów, uznając go za sprawcę operacji tłumienia frekwencji. Na Tajwanie, w tygodniach poprzedzających styczniowe wybory prezydenckie, w platformach Line, Facebook, X i TikTok krążył zmontowany film przedstawiający Lai Ching-tego, kandydata Demokratycznej Partii Postępowej, wypowiadającego kwestie sugerujące istnienie nieślubnego dziecka i skandalu obyczajowego[2]. W Argentynie sztaby Sergia Massy i Javiera Milei publikowały *deepfake*'i i grafiki generowane przez AI z taką intensywnością, że amerykańscy komentatorzy określili tę kampanię mianem „pierwszych wyborów ery sztucznej inteligencji"[3].

Sekwencja tych wydarzeń stanowi punkt wyjścia do dalszych rozważań. Niniejszy artykuł stawia sobie dwa cele. Pierwszym jest zestawienie głównych nurtów badań akademickich nad wpływem generatywnej AI na dezinformację polityczną. Drugim: krytyczne rozważenie hipotezy, iż dopiero wraz z rozpowszechnieniem dużych modeli językowych rozpoczęła się „nowa era dezinformacji". Analiza nie ogranicza się do rejestru zagrożeń. Rozważane są także mechanizmy społeczne i psychologiczne, dzięki którym określone treści zyskują wiarygodność, oraz strategie przeciwdziałania: regulacyjne, edukacyjne i technologiczne.

Praca ma charakter przeglądu narracyjnego uzupełnionego o analizę wybranych studiów przypadków. Materiał źródłowy pozyskano z dwóch grup: recenzowanych artykułów, monografii oraz raportów instytucjonalnych sygnowanych przez uznane ośrodki analityczne (Oxford Internet Institute, Center for Security and Emerging Technology, Council of Europe, EU DisinfoLab, NATO StratCom, SGDSN).

Selekcja studiów przypadków opierała się na trzech kryteriach: aktualności (lata 2023–2025), różnorodności geograficznej (Europa Środkowa, Ameryka Północna, Ameryka Południowa, Azja Wschodnia) oraz udokumentowaniu w li-

[1] L. de Nadal, P. Jančárik, *Beyond the deepfake hype: AI, democracy, and „the Slovak case"*, „Harvard Kennedy School (HKS) Misinformation Review" 2024, t. 5, nr 4, https://eprints.gla.ac.uk/333811/1/333811.pdf [dostęp: 12.07.2026].

[2] A. H.-E. Wang, *AI-generated disinformation in the 2024 Taiwan presidential election*, Friedrich Naumann Foundation for Freedom, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf [dostęp: 12.07.2026].

[3] J. Nicas, L. C. Herrera, *Is Argentina the First A.I. Election?*, „The New York Times", 14 listopada 2023.

teraturze fachowej. Odrzucono przypadki znane głównie z doniesień prasowych bez pogłębionej analizy akademickiej.

1. Konceptualizacja: między misinformacją, dezinformacją a nowym repertuarem AI

Zjawiska omawiane w niniejszym artykule funkcjonują w literaturze pod kilkoma pokrewnymi nazwami. Ustalone rozróżnienia porządkują analizę.

Powszechnie przyjmowana typologia pochodzi z raportu Rady Europy autorstwa Claire Wardle i Hosseina Derakhshana[4]. Autorzy proponują pojęcie „nieporządku informacyjnego" (*information disorder*) obejmującego trzy zjawiska:

- misinformację (*misinformation*): fałszywe treści udostępniane bez intencji szkodzenia, często w wyniku pomyłki lub emocjonalnej reakcji;
- dezinformację (*disinformation*): fałszywe treści wytwarzane i rozpowszechniane celowo dla wyrządzenia szkody określonemu podmiotowi (państwu, partii, grupie społecznej);
- malinformację (*malinformation*): treści prawdziwe wyrwane z kontekstu i użyte instrumentalnie w celu wyrządzenia szkody (np. wyciek prywatnych zdjęć)[5].

Zaletą tej triady jest oddzielenie intencji od prawdziwości. Klasyczna „propaganda" z okresu zimnowojennego mieściła się głównie w dezinformacji, tymczasem współczesny ekosystem obejmuje wszystkie trzy kategorie i pozwala aktorom łatwo się między nimi przemieszczać. Wardle i Derakhshan podkreślają także rolę „zanieczyszczenia informacyjnego" (*information pollution*) rozumianego jako systemowe zjawisko, nie zaś jako suma pojedynczych fałszów.

Termin *deepfake* powstał natomiast w 2017 roku jako połączenie *deep learning* i *fake*. Chesney i Citron definiują *deepfake*'i jako „hiperrealistyczne cyfrowe fałszerstwa obrazu, dźwięku lub wideo, tworzone przy użyciu głębokich sieci neuronowych", zdolne przekonująco podmienić twarz, głos i mimikę osoby publicznej[6]. Autorzy przewidzieli już wtedy skutki tej technologii dla prywatności, procedur sądowych, dyskursu politycznego i bezpieczeństwa narodowego, a zbiór ich analiz stał się podstawą dla późniejszych regulacji.

Obok *deepfake*'ów pojawiła się kategoria *cheapfake* manipulacja tanimi, dobrze znanymi środkami (przyspieszenie, zwolnienie, wybiórcze cięcie, dublaż

[4] C. Wardle, H. Derakhshan, *Information disorder: Toward an interdisciplinary framework for research and policy making* (DGI(09(2017), Council of Europe, https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html [dostęp: 12.07.2026].

[5] Tamże.

[6] R. Chesney, D. K. Citron, *Deep fakes: A looming challenge for privacy, democracy, and national security,* „California Law Review" 2017, t. 107, nr 6, s. 1753–1820.

lektorski). *Cheapfake* nie wymaga sztucznej inteligencji, natomiast wywołuje analogiczne skutki poznawcze[7]. Rozgraniczenie tych zjawisk ma znaczenie polityczne: bardzo często kampania nazywana *deepfake*'em opiera się w istocie na *cheapfake*'u, co utrudnia diagnozę i osłabia zaufanie do *fact-checkingu*.

Termin *generatywna sztuczna inteligencja* obejmuje modele, które produkują nowe treści (tekst, obraz, wideo, dźwięk) na podstawie danych uczących. Kreps, McCain i Brundage już w eksperymentach z GPT-2 pokazali, że laboratoryjnie generowane teksty imitujące artykuły prasowe zyskiwały wiarygodność zbliżoną do materiałów autentycznych[8]. Ich ustalenia okazały się prekursorskie wobec późniejszej fali GPT-3 i GPT-4. Kluczową obserwacją nie jest fakt, że modele „potrafią kłamać", tylko że kosztowo dezinformacja przesunęła się z aktywności wymagającej sztabu redakcyjnego na aktywność jednoosobową i praktycznie darmową[9].

2. Badania nad *fake news* i mediami społecznościowymi

Współczesna fala badań rozpoczęła się po wyborach prezydenckich w Stanach Zjednoczonych w 2016 roku. Allcott i Gentzkow obliczyli, że tuż przed listopadowym głosowaniem najbardziej wiralowe „historie *fake news*: sprzyjające Trumpowi były udostępniane 30 milionów razy, a te sprzyjające Clinton – 8 milionów razy"[10]. Autorzy zachowali ostrożność w wnioskach o wpływie na wynik wyborów, wskazywali jednak na wyraźną asymetrię ekspozycji[11]. W tym samym okresie Bessi i Ferrara wykazali, że około 20% dyskusji politycznej na Twitterze w okresie kampanii pochodziło od kont automatycznych. Późniejsze prace Ferrary rozszerzyły diagnozę na kolejne wybory (Brazylia 2018, Meksyk 2018, USA 2020) i wprowadziły podstawy klasyfikacji „propagandy obliczeniowej" (*computational propaganda*)[12].

7 B. Paris, J. Donovan, *Deepfakes and cheap fakes: The manipulation of audio and visual evidence,* https://datasociety.net/research-library/deepfakes-and-cheap-fakes-the-manipulation-of-audio-and-visual-evidence/ [dostęp: 12.07.2026].

8 S. Kreps, R. M. McCain, M. Brundage, *All the news that's fit to fabricate: AI-generated text as a tool of media misinformation,* „Journal of Experimental Political Science" 2020, t. 9, nr 1, s. 104–117.

9 B. Buchanan, A. Lohn, M. Musser K. Sedova, *Truth, lies, and automation: How language models could change disinformation,* Center for Security and Emerging Technology, Georgetown University, https://cset.georgetown.edu/publication/truth-lies-and-automation/ [dostęp: 12.07.2026].

10 H. Allcott, M. Gentzkow, *Social media and fake news in the 2016 election*, „Journal of Economic Perspective" 2017, t. 31, nr 2, s. 211–236.

11 Tamże.

12 A. Bessi, E. Ferrara, *Social bots distort the 2016 U.S. presidential election online discussion,* „First Monday" 2016, t. 21, nr 11, https://firstmonday.org/ojs/index.php/fm/article/view/7090 [dostęp: 12.07.2026].

Vosoughi, Roy i Aral opublikowali w „Science" analizę 126 tys. „historii" na Twitterze z lat 2006–2017. Zebrany materiał obejmował około 4,5 mln tweetów od 3 milionów użytkowników. Fałszywe wiadomości docierały do 1 500 osób średnio sześciokrotnie szybciej niż prawdziwe, a największą przewagę zyskiwały informacje polityczne. Autorzy tłumaczyli różnicę „nowością" fałszów oraz emocjami, które wywoływały: strach, obrzydzenie, zaskoczenie. Wynik ten pozostaje jednym z najczęściej cytowanych argumentów za tezą o strukturalnej podatności mediów społecznościowych na rozpowszechnianie fałszu[13].

Syntetyzując dorobek psychologii poznawczej wskazuje się, że łatwiej niż motywacja polityczna wiarę w *fake news* tłumaczy „leniwe myślenie": przetwarzanie automatyczne, brak refleksji nad zawartością nagłówka, wysoka podatność na treści potwierdzające istniejące przekonania[14]. Warto także wspomnieć o fenomenie „*backfire effect*": w niektórych warunkach korekta fałszywej informacji wzmacnia u odbiorcy pierwotne, błędne przekonanie. Późniejsze próby replikacji podważyły uniwersalność tego zjawiska, ale sam mechanizm motywowanego przetwarzania informacji pozostał ważnym punktem odniesienia dla polityk publicznych[15]. Nowsze badania wskazały, że dezinformacja podważa nie tyle „prawdziwość poszczególnych sądów", ile „epistemiczną integralność demokracji". Argumentacja odchodzi od wąskiej dyskusji o poszczególnych treściach na rzecz analizy systemowej: kiedy obywatele przestają wierzyć procedurom, ekspertom i sobie nawzajem, każde głosowanie staje się polem sporu o rzeczywistość, nie o politykę[16].

Model biznesowy głównych platform mediów społecznościowych opiera się na maksymalizacji zaangażowania, a maksymalizacja zaangażowania promuje treści emocjonalne, polaryzujące i sensacyjne. Dezinformacja w tej perspektywie nie jest anomalią, lecz naturalnym produktem architektury systemu. Użytkownicy spotykają się z większą różnorodnością źródeł niż zwolennicy „tradycyjnych" mediów. Sam mechanizm algorytmicznej kuracji treści nie budzi jednak wątpliwości: to platformy, nie użytkownicy, decydują, jakie treści zyskają zasięg[17].

Rozwój ChatGPT i modeli obrazowych po 2022 roku wywołał dwie kontrastujące narracje w literaturze akademickiej. Pierwsza narracja podkreśla „skok

[13] S. Vosoughi, D. Roy, S. Aral, *The spread of true and false news online*, „Science" 2018, t. 359, nr 6380, s. 1146–1151.

[14] G. Pennycook, D. G. Rand, *The psychology of fake news*, „Trends in Cognitive Sciences" 2021, t. 25, nr 5, s. 388–402.

[15] B. Nyhan, J. Reifler, *When corrections fail: The persistence of political misperceptions*, „Political Behavior" 2010, t. 32, nr 2, s. 303–330.

[16] S. Lewandowsky, S. van der Linden, *Countering misinformation and fake news through inoculation and prebunking*, „European Review of Social Psychology" 2021, t. 32, nr 2, s. 348–384.

[17] N. Persily, J. A. Tucker (Eds.), *Social media and democracy: The state of the field, prospects for reform,* Cambridge 2020.

jakościowy". Modele językowe klasy GPT-3 mogą przejąć kluczowe funkcje operacji dezinformacyjnych: personalizację przekazu, wielojęzyczną produkcję, ciągłe generowanie nowych wariantów treści. Ich eksperymenty wykazały, że GPT-3, poddany minimalnemu tuningowi, produkuje przekonujące teksty wspierające narracje prorosyjskie oraz teorie spiskowe[18]. Druga narracja przekonuje, że panika jest przedwczesna. Generatywna AI zwiększa podaż fałszów, ale to nie podaż stanowi ograniczenie systemu. Ograniczeniem jest popyt: konsumpcja treści dezinformacyjnych i tak koncentruje się w wąskich grupach, a większość odbiorców czerpie informacje z mediów mainstreamowych. W tej perspektywie skutki AI w krótkim horyzoncie będą marginalne[19].

3. Sztuczna inteligencja a ekosystem dezinformacji

Najczęściej wskazuje się na cztery wymiary, w których duże modele językowe mogą zmienić operacje dezinformacyjne: skalę, personalizację, wielojęzyczność i szybkość reakcji[20]. Poniższa analiza wykorzystuje ten podział, wzbogacając go o dwa dodatkowe wymiary: produkcję *deepfake*'ów audiowizualnych oraz „dywidendę kłamcy".

a) Skala produkcji

Klasyczna „farma trolli" wymaga zatrudnienia setek osób produkujących komentarze, artykuły i memy. Petersburska Agencja Badań Internetowych, ujawniona w śledztwie Muellera, zatrudniała w szczycie działalności ponad 400 osób. Model GPT-4, uruchamiany na standardowej infrastrukturze chmurowej, wytwarza analogiczną objętość tekstu w ciągu kilku minut. Kreps w kontrolowanym eksperymencie pokazał, że amerykańscy respondenci nie odróżniali tekstów wygenerowanych przez GPT-2 od tekstów redakcyjnych „New York Timesa" w stopniu istotnym statystycznie[21]. Uwzględniając późniejszą ewolucję modeli, można ostrożnie przyjąć, że koszt marginalny „wiarygodnego artykułu propagandowego" zmierza obecnie do zera.

b) Personalizacja i mikrotargetowanie

Modele generatywne pozwalają dostosowywać przekaz do domniemanego profilu odbiorcy. Mikrotargetowanie oparte na uczeniu maszynowym prowadzi do

[18] B. Buchanan, A. Lohn, M. Musser K. Sedova, *Truth, lies, and automation*.

[19] F. M. Simon, S. Altay H. Mercier, *Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown*, „Harvard Kennedy School (HKS) Misinformation Review" 2023, t. 4, nr 5.

[20] B. Buchanan, A. Lohn, M. Musser K. Sedova, *Truth, lies, and automation*.

[21] S. Kreps, R. M. McCain, M. Brundage, *All the news that's fit to fabricate*.

sytuacji, w której wyborcy „otrzymują spersonalizowane do ich poglądów i opinii komunikaty", co utrudnia rzetelną weryfikację i wzmacnia „bańkę informacyjną". Efekt ten łączy technologię AI z architekturą rekomendacji platform: dopasowany przekaz otrzymuje priorytet algorytmiczny, co dodatkowo zwiększa jego zasięg[22].

c) Wielojęzyczność

Jednym z ograniczeń tradycyjnych operacji wpływu była bariera językowa. Rosyjskojęzyczna farma trolli wytwarzała słabe teksty po angielsku, chińska: słabe teksty po hiszpańsku. Modele klasy GPT-4 lub Claude wygenerują poprawne stylistycznie teksty w kilkudziesięciu językach jednocześnie, dostosowując się do lokalnych rejestrów. W rezultacie państwa niebędące dotąd typowymi celami operacji rosyjskich czy chińskich (np. Peru, Węgry, Portugalia) mogą zyskać znaczenie strategiczne właśnie dlatego, że wcześniejsza bariera została zniesiona.

d) Szybkość reakcji i personifikacja botów

Ferrara w podsumowaniu badań nad botami politycznymi opisał moment przełomu: przejście od kont automatycznych publikujących z góry przygotowane teksty do „botów konwersacyjnych" reagujących w czasie rzeczywistym. Wraz z upowszechnieniem LLM bariera odróżnienia bota od człowieka drastycznie się zmniejszyła[23].

e) *Deepfake* audiowizualny

Można wskazać cztery zagrożenia płynące z rozwoju *deepfake*'ów: destrukcyjne skutki dla prywatności, zagrożenie dla procedur sądowych, ryzyko destabilizacji politycznej oraz stępienie zdolności odbiorców do odróżniania prawdy od fałszu[24]. Wraz z upowszechnieniem takich narzędzi jak ElevenLabs (klonowanie głosu z próbki 30 sekund) i modeli obrazowych (Stable Diffusion, Midjourney), technologia stała się dostępna dla użytkowników indywidualnych, nie tylko dla wyspecjalizowanych aktorów państwowych. Przypadek *robocalli* z głosem Bidena w New Hampshire jest tego przykładem: pojedynczy konsultant polityczny, korzystający z komercyjnie dostępnego narzędzia, doprowadził do operacji tłumienia frekwencji na skalę stanu.

[22] K. Mroczka, B. Biderma, *Fake newsy i deepfaki w kampanii wyborczej w Polsce w 2023 roku – case studies*, „Studia Politologiczne" 2025, nr 74, s. 23–52.

[23] E. Ferrara, *Bots, elections, and social media: A brief overview*, s. 95–114, w: K. Shu, S. Wang, D. Lee, H. Liu (Eds.), Disinformation, misinformation, and fake news in social media, Springer, 2020.

[24] R. Chesney, D. K. Citron, *Deep fakes: A looming challenge for privacy*.

f) Dywidenda kłamcy

Pojęcie „*liar's dividend*" wprowadzili Chesney i Citron: w środowisku, w którym każdy materiał może być syntetyczny, aktorzy polityczni zyskują możliwość zaprzeczania autentycznym nagraniom pod pretekstem ich rzekomej fałszywości. W dłuższej perspektywie zjawisko to podważa samą kategorię „dowodu": nawet nagranie autentyczne nie jest w stanie zamknąć debaty, jeżeli można prawdopodobnie zasugerować, że pochodzi z AI[25].

4. Studia przypadków

4.1. Słowacja 2023: pierwszy „test wyborczy" ery AI

28 września 2023 roku, dwa dni przed głosowaniem, w mediach społecznościowych pojawiło się nagranie audio, na którym rzekomo słychać było Michala Šimečkę omawiającego z dziennikarką *Denníka N*, Moniką Tódovą, plan kupowania głosów w społeczności romskiej. Materiał krążył podczas obowiązującej w Słowacji 48-godzinnej ciszy wyborczej, co ograniczyło możliwości dementowania w tradycyjnych mediach. Analizy techniczne przeprowadzone przez laboratoria *fact-checkingowe* (m.in. Central European Digital Media Observatory) wykazały cechy syntezy głosu, natomiast do dziś nie ustalono jednoznacznie sprawcy. Prokuratura słowacka wszczynała trzykrotnie postępowania, wszystkie umarzano z powodu braku dowodów.

Przypadek słowacki jest często określany jako „prekursor" szerszego zjawiska: operacja miała ograniczoną skalę techniczną, ale otworzyła katalog scenariuszy powielanych następnie w kolejnych krajach. Kluczową obserwacją była asymetria czasowa: sfabrykowanie i wypuszczenie *deepfake*'a zajęło godziny, jego rzetelne rozbicie – dni i tygodnie[26].

De Nadal i Jančárik zajmują natomiast stanowisko krytyczne wobec narracji o „pierwszych wyborach zrewolucjonizowanych przez AI". Ich zdaniem zwycięstwo Fica należy interpretować w kontekście długoterminowych procesów: (a) spadku zaufania do mediów mainstreamowych (Słowacja plasowała się w 2023 roku w dolnej połowie rankingu Reuters Digital News Report); (b) rozczarowania kolejnymi rządami; (c) słowiańskich związków kulturowych z Rosji; (d) narracji „zmęczenia Ukrainą" wzmocnionej presją energetyczną; (e) roli komunikatora Telegram jako głównego kanału propagandy prorosyjskiej[27].

[25] Tamże.

[26] M. Bohacek, *Slovakia as the precursor to deepfake-enabled election interference: Lessons learned and pathways forward*, https://workshop-proceedings.icwsm.org/abstract.php?id=2024_67 [dostęp: 12.07.2026].

[27] L. de Nadal, P. Jančárik, *Beyond the deepfake hype: AI, democracy*.

4.2. Stany Zjednoczone 2024: *robocall* Bidena i szerszy krajobraz

W nocy 22 stycznia 2024 roku, dwie doby przed prawyborami demokratycznymi w New Hampshire, tysiące zarejestrowanych wyborców Demokratów odebrało nagrania telefoniczne. Głos, przypominający głos prezydenta Bidena, rozpoczynał się od charakterystycznego „What a bunch of malarkey" i wzywał do niegłosowania w prawyborach: „save your vote for November". Nagrania powstały przy użyciu narzędzia ElevenLabs, po sklonowaniu głosu z próbki wystąpień publicznych. FCC uznała później organizatorów za odpowiedzialnych za nielegalny *spoofing* numerów (podszywanie się pod numery lokalnych polityków demokratycznych) oraz naruszenie Telephone Consumer Protection Act. Kara nałożona na Steve'a Kramera – konsultanta demokratycznego pracującego wcześniej dla sztabów – wyniosła 6 mln USD.

Raport Brookings Institution wskazuje, że w kampanii prezydenckiej 2024 roku udokumentowano dziesiątki incydentów treści syntetycznych: od zmanipulowanych obrazów Kamali Harris, Trumpa, przez fałszywe filmy z Muskiem wypowiadającym poparcie polityczne, po fabrykowane „lekarskie diagnozy" kandydatów krążące w wąskich niszach społecznościowych. Wpływ tych operacji na wynik pozostał ograniczony: wyborcy traktowali podejrzane materiały ze zdrowym sceptycyzmem, a ekosystem *fact-checkingowy* zdołał w większości przypadków szybko dementować fałsze[28].

W badaniu dotyczącym pandemii Kreps i Kriner pokazali, że amerykański wyborca sytuuje AI-dezinformację wysoko na liście postrzeganych zagrożeń dla demokracji. Autorzy podnoszą, że sama „percepcja zagrożenia" może okazać się destrukcyjna niezależnie od realnego wpływu poszczególnych operacji: jeśli obywatele przestają wierzyć, że mogą odróżnić prawdę od fałszu, tracą motywację do informowania się i mobilizacji politycznej[29].

4.3. Polska 2023: spot z „głosem Tuska" i szeroki wachlarz manipulacji

W polskiej kampanii parlamentarnej 2023 roku odnotowano kilka wyrazistych incydentów. Najgłośniejszym była publikacja spotu Prawa i Sprawiedliwości „Tusk nie zasługuje na kolejną szansę", w którym część wypowiedzi rzekomo pochodzących z autentycznych źródeł została w istocie zsyntetyzowana lub wyrwana z kontekstu. Wcześniej, w sierpniu 2023 roku, poseł Solidarnej Polski, Piotr Sak, opublikował materiał, w którym „głos Tuska" – wygenerowany przez syntezę – odczytywał treść e-maili wysłanych przez byłego premiera.

[28] D. M. West, *How disinformation defined the 2024 election narrative,* https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/ [dostęp: 12.07.2026].

[29] S. E. Kreps. D. L. Kriner, *Assessing misinformation recall and accuracy perceptions: Evidence from the COVID-19 pandemic,* „Harvard Kennedy School (HKS) Misinformation Review" 2023, t. 4, nr 5.

Wykorzystanie treści zmanipulowanych i syntetycznych wpisywało się w szerszą logikę komunikacji politycznej opartej na emocjonalizacji przekazu i delegitymizowaniu przeciwnika. Przykłady materiałów z udziałem syntetyzowanego głosu czy selektywnie montowanych wypowiedzi nie były jedynie incydentami technologicznymi, lecz elementami strategii perswazyjnej, w której granica między autentycznością a manipulacją została celowo rozmyta. Kluczowe znaczenie miała tu nie tyle sama obecność technologii AI, ile sposób jej użycia – jako narzędzia wzmacniającego istniejące narracje polityczne oraz upraszczającego przekaz poprzez jego personalizację i emocjonalne nasycenie[30].

Istotnym aspektem była również dynamika rozprzestrzeniania się takich treści. Materiały o charakterze manipulacyjnym funkcjonowały przede wszystkim w środowisku mediów społecznościowych, gdzie ich zasięg był wzmacniany przez algorytmy premiujące zaangażowanie użytkowników. Oznacza to, że nawet relatywnie proste technicznie manipulacje mogły osiągać szeroką widoczność, jeśli wpisywały się w istniejące podziały polityczne i odpowiadały na oczekiwania określonych grup odbiorców. W tym sensie skuteczność dezinformacji nie wynikała z jej jakości technologicznej, lecz z dopasowania do kontekstu społecznego i politycznego[31].

Równolegle kampania ujawniła rosnącą rolę infrastruktury przeciwdziałania dezinformacji, jednak jej działanie miało charakter reaktywny. Mechanizmy weryfikacyjne, choć istotne, funkcjonowały z opóźnieniem względem tempa dystrybucji treści, co ograniczało ich skuteczność w kształtowaniu bieżącej percepcji wyborców. W praktyce oznaczało to, że sprostowania docierały do odbiorców później niż pierwotne, często fałszywe przekazy[32].

Całość zjawiska należy interpretować jako przejaw pogłębiającej się transformacji komunikacji politycznej, w której technologie cyfrowe – w tym generatywna sztuczna inteligencja – pełnią funkcję katalizatora, a nie głównego źródła problemu. Polska kampania z 2023 roku pokazuje, że nawet ograniczone użycie zaawansowanych narzędzi może mieć znaczący wpływ na jakość debaty publicznej, jeśli trafia na podatny grunt w postaci silnej polaryzacji i niskiego poziomu zaufania do instytucji.

4.4. Tajwan 2024: kampania w cieniu chińskich operacji wpływu

Wybory prezydenckie na Tajwanie w styczniu 2024 roku odbyły się w warunkach intensywnej presji informacyjnej z ChRL. Można było w ich trakcie zaobserwować następujące typy działań:

[30] K. Mroczka, B. Biderma, *Fake newsy i deepfaki w kampanii*.

[31] D. Piontek, *Wyborcy i media. Ocena działań mediów w okresie polskiej parlamentarnej kampanii wyborczej 2023 roku,* „Przegląd Politologiczny" 2025, t. 30, nr 1.

[32] H. Świeboda, *Generatywna sztuczna inteligencja wyzwaniem dla bezpieczeństwa społecznego i państwa,* „Maritime Security Yearbook" 2025, t. 19.

Modyfikowane przez AI wideo. Po publikacji spotu wyborczego Lai Ching-tego pt. *On the Road* pojawiła się przerobiona wersja z syntetycznie zmienioną synchronizacją ust i głosem, sugerująca istnienie nieślubnego dziecka. Materiał krążył w formie masowo udostępnianej na Facebooku, X, TikToku i w komunikatorze Line.

Fikcyjne konta z twarzami generatywnymi. Grupy kont z profilowymi zdjęciami tworzonymi przez modele obrazowe: taniej, trudniej wykrywalne, przypominające lokalnych mieszkańców.

Wirtualni prezenterzy YouTube. Ponad 20 nowych kanałów prezentowało materiały syntetycznych „prezenterów" odczytujących artykuły chińskich mediów państwowych[33].

Mimo intensywności ataków ich wpływ na wynik był ograniczony. Tajwan wypracował w poprzednich latach infrastrukturę odporności: Taiwan FactCheck Center, IORG, Taiwan AI Labs, CoFacts. Kluczowa okazała się także kultura obywatelska – zaangażowanie społeczne w weryfikację treści zbliżało się do stanu „rozproszonego dziennikarstwa śledczego"[34].

4.5. Argentyna 2023: „pierwsze wybory ery AI"

Kampania prezydencka w Argentynie w 2023 roku stała się jednym z pierwszych przypadków szerokiego i jawnego wykorzystania generatywnej sztucznej inteligencji przez konkurujące ze sobą sztaby wyborcze. Zarówno obóz Sergia Massy, jak i Javiera Milei wykorzystywał AI do produkcji treści wizualnych i audiowizualnych – od stylizowanych plakatów propagandowych po materiały wideo i memy – które służyły zarówno autopromocji, jak i delegitymizacji przeciwnika. Charakterystyczne było tu nie tyle ukrywanie manipulacji, ile jej normalizacja jako elementu repertuaru komunikacyjnego[35].

Z perspektywy analitycznej kluczowe znaczenie ma zróżnicowanie funkcji tych treści. Część materiałów miała charakter dezinformacyjny – jak *deepfake* przedstawiający Massę w sytuacji kompromitującej – i opierała się na wzmacnianiu już istniejących narracji i plotek, co zwiększało ich wiarygodność wśród odbiorców. Równolegle funkcjonowały treści jawnie fikcyjne, wykorzystujące estetykę kultury popularnej i hiperbolę, które nie miały na celu bezpośredniego

[33] A. H.-E. Wang, *AI-generated disinformation in the 2024 Taiwan presidential*

[34] C.-L. Hung, W.-C. Fu, C.-C. *Liu, AI disinformation attacks and Taiwan's responses during the 2024 presidential election*, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf [dostęp: 12.07.2026].

[35] OECD.AI., *AI-generated deepfakes manipulate Argentine presidential election*. OECD AI Incidents and Hazards Monitor, https://oecd.ai/en/incidents/2023-11-14-1e5d [dostęp: 12.07.2026].

wprowadzenia w błąd, lecz budowanie emocjonalnych skojarzeń i uproszczonych ocen politycznych[36].

Jednocześnie dostępne analizy wskazują, że rola AI w Argentynie była wtórna wobec czynników strukturalnych. O wyniku wyborów zdecydowały przede wszystkim kryzys gospodarczy, wysoka inflacja oraz silna mobilizacja elektoratu antyestablishmentowego. Generatywna AI działała raczej jako „wzmacniacz" istniejących napięć – obniżając koszty produkcji treści, zwiększając ich atrakcyjność wizualną i dopasowanie do logiki mediów społecznościowych – niż jako czynnik samodzielnie determinujący wynik wyborczy.

4.6. Doppelganger: rosyjska operacja podszywania się pod media

„Doppelganger" to nazwa nadana przez EU DisinfoLab operacji wpływu prowadzonej od 2022 roku przez rosyjskie podmioty (m.in. Structura National Technologies i Social Design Agency, oba objęte sankcjami UE i USA). Operacja polega na tworzeniu domen imitujących europejskie i amerykańskie serwisy informacyjne („Le Monde", „Der Spiegel", „Bild", „The Washington Post", „20 Minutes") oraz publikowaniu na nich fabrykowanych artykułów, dystrybuowanych następnie przez konta w mediach społecznościowych. Raport SGDSN z 2025 roku wskazuje, że tylko we Francji odnotowano w 2024 roku ponad 300 zestawów treści wpisujących się w Doppelganger[37].

W kontekście AI operacja pokazuje inne oblicze problemu: nie tyle spektakularny *deepfake*, ile mozolne, powtarzalne, algorytmicznie wspomagane działania, których zasięg mierzy się nie w milionach, ale w setkach tysięcy interakcji z „wąską", zbudowaną kulturowo grupą odbiorców. Mimo blokad platform Doppelganger regeneruje się przez uruchamianie nowych domen szybciej, niż zdołano je usuwać: aktorzy AI-wspomagani wypracowali cykl życia „kampanii tydzień"[38].

Istotnym elementem skuteczności operacji „Doppelganger" jest także jej adaptacyjność i niski koszt reprodukcji. Tworzenie kolejnych domen imitujących znane media nie wymaga dużych zasobów technologicznych, natomiast dzięki automatyzacji (w tym narzędziom AI wspierającym generowanie i modyfikację treści) możliwe jest szybkie dostosowywanie przekazu do bieżących wydarzeń politycznych oraz lokalnych kontekstów. Operacja działa więc w logice „ciągłej

[36] A. Slimovich, *Desinformación e inteligencia artificial en las redes sociales en la Argentina*, w: D. H. Cabrera-Altieri, G. López-García & E. Campos-Domínguez (red.), *Perturbaciones informativas. Desinformación y mediatización digital*. Comunicación Social Ediciones y Publicaciones, 2025.

[37] *Summary of the Information Threat to the Paris 2024 Olympic and Paralympic* Games, https://www.sgdsn.gouv.fr/files/files/Publications/20240919_NP_SGDSN_VIGINUM_Summary%20information%20threat%20Paris2024Games_EN_0.pdf [dostęp: 12.07.2026].

[38] M. G. Sessa, R. Miguel, *The Doppelganger case: Assessment of platform regulation on the EU disinformation environment,* NATO Strategic Communications Centre of Excellence 2024.

obecności", a nie jednorazowego uderzenia: zamiast spektakularnych kampanii obserwujemy rozproszoną, trudną do uchwycenia aktywność, która wykorzystuje luki w moderacji platform i asymetrię między tempem publikacji a tempem reakcji instytucji. W rezultacie nawet częściowo zdezaktywowane sieci dezinformacyjne zachowują zdolność oddziaływania, ponieważ ich struktura nie jest scentralizowana, lecz opiera się na powielalnych modułach – domenach, kontach i narracjach – które mogą być odtwarzane niemal natychmiast po usunięciu.

5. Konsekwencje społeczno-polityczne

a) Erozja zaufania publicznego

Sama ekspozycja na *fake news* wiąże się z niższym zaufaniem do mediów i wyższym zaufaniem do rządu wtedy, gdy „nasza strona" sprawuje władzę. Mechanizm jest asymetryczny: dezinformacja niszczy nie zaufanie w ogóle, ale zaufanie do tych instytucji, które odbiorca postrzega jako „obce" ideologicznie. W dłuższej perspektywie prowadzi to do struktury, w której obywatele mają zaufanie wyłącznie do własnej „bańki"[39].

b) Polaryzacja polityczna

Media społecznościowe wzmacniają polaryzację przez połączenie „bańki filtracyjnej" z „efektem tożsamościowym": użytkownicy wystawieni na skrajne treści odmiennego obozu nie „miękną", ale utwardzają się w swoich pozycjach. Badania eksperymentalne na Twitterze pokazały, że sam kontakt z odmiennym punktem widzenia w warunkach algorytmicznie moderowanego dyskursu prowadzi do wzrostu polaryzacji, nie zaś do jej redukcji. W kontekście generatywnej AI mechanizm ten zyskuje nową dynamikę: masowe wytwarzanie personalizowanych treści potwierdzających polaryzowane postawy staje się operacją niskokosztową[40].

c) Wpływ na wybory i procedury demokratyczne

Uporczywe pytanie „czy AI zdecyduje o wyniku wyborów?" jest w istocie źle sformułowane. Wybory rzadko rozstrzyga pojedyncze zdarzenie; częściej decyduje kumulatywna atmosfera epistemiczna. *Deepfake* Šimečki nie zdecydował o zwycięstwie Fica, ale wpisał się w mozaikę czynników, które łącznie odbiły się na wyniku[41]. *Robocall* Bidena nie zdecydował o wyniku prawyborów w New

[39] K. Ognyanova, D. Lazer, R. E. Robertson, C. Wilson, *Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power*, „Harvard Kennedy School (HKS) Misinformation Review" 2020, t. 1, nr 4.

[40] C. A. Bail, *Breaking the social media prism: How to make our platforms less polarizing*, Princeton University Press 2021.

[41] L. de Nadal, P. Jančárik, *Beyond the deepfake hype: AI, democracy*.

Hampshire, ale ustanowił precedens, którego kolejne kampanie już nie mogły zignorować. Właściwym pytaniem staje się nie „czy wybory zostały zdecydowane", tylko „czy procedura wyborcza zachowała integralność jako proces zbiorowego rozstrzygania"[42].

d) Dywidenda kłamcy i kryzys dowodu

Najgroźniejszym efektem *deepfake*'ów nie są sfabrykowane materiały same w sobie, tylko możliwość dyskredytacji materiałów autentycznych. W środowisku, w którym każde wideo może być syntetyczne, aktor polityczny nakrywany na nagraniu ma prostą linię obrony: „to *deepfake*". Możliwość ta była już wykorzystywana. W 2018 roku prezydent Gabonu, Ali Bongo, po epizodzie zdrowotnym opublikował wideo, którego autentyczność podważyli opozycjoniści; niepewność ta towarzyszyła nieudanej próbie zamachu stanu. Trump w kampanii 2024 roku sugerował, że pewne kompromitujące go materiały były „AI-generated". Dywidenda kłamcy podważa nie tylko poszczególne dowody, ale całą praktykę argumentacji opartej na materiale audiowizualnym[43].

e) Efekt „taniej propagandy" dla reżimów autorytarnych

Już przed erą dużych modeli językowych badacze z Oxford Internet Institute, Samantha Bradshaw i Philip N. Howard, udokumentowali zorganizowane kampanie manipulacji w mediach społecznościowych prowadzone w 70 państwach.

Generatywna AI radykalnie obniża próg wejścia. Reżim niewielki finansowo i technologicznie może dziś prowadzić operacje wpływu o skali, która wcześniej była zarezerwowana dla Rosji, Chin i USA. W rezultacie „wyścig zbrojeń" w domenie informacyjnej ulegnie demokratyzacji, co paradoksalnie może zaszkodzić stabilności demokracji[44].

6. Strategie przeciwdziałania

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689, przyjęte 13 czerwca 2024 roku, znane jako AI Act, wprowadza obowiązki transparentności dla dostawców i użytkowników systemów AI produkujących treści syntetyczne. Artykuł 50 nakłada obowiązek oznakowania *deepfake*'ów przeznaczonych do publikacji oraz obowiązek osadzania „maszynowo czytelnych" oznaczeń w treściach generowanych przez AI. AI Act stanowi pierwszy horyzontalny akt prawny na

[42] F. M. Simon, S. Altay, H. Mercier, *Misinformation reloaded? Fears about the impact of generative AI*.

[43] R. Chesney, D. K. Citron, *Deep fakes: A looming challenge for privacy*.

[44] S. Bradshaw, P. N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*, „Working Paper" 2019, nr 3.

świecie, który wprost adresuje *deepfake*'i, natomiast jego skuteczność zależy od faktycznego egzekwowania (obowiązki wchodzą etapami do sierpnia 2026 roku)[45]. Rozporządzenie 2022/2065 (Digital Services Act), obowiązujące od 2024 roku, ustanawia z kolei ramy odpowiedzialności platform. Bardzo duże platformy internetowe (VLOP) i wyszukiwarki (VLOSE) podlegają obowiązkowi rocznej oceny ryzyk systemowych, w tym ryzyka „negatywnego wpływu na dyskurs obywatelski i procesy wyborcze". Mimo istniejących uprawnień DSA nie zdołał na razie zatrzymać kampanii rosyjskiej: platformy wykonują polecenia usunięcia pojedynczych treści, ale ekosystem regeneruje się szybciej, niż podejmowane są działania[46].

Ważnym komponentem jest edukacja. Lewandowsky i van der Linden rozwinęli koncepcję „inokulacji psychologicznej" (*prebunking*): ekspozycja odbiorców na słabe wersje technik manipulacji buduje odporność analogicznie do szczepionki biologicznej. Badania eksperymentalne prowadzone we współpracy z Google i YouTube pokazały statystycznie istotne, choć umiarkowane efekty ochronne po ekspozycji na krótkie animacje edukacyjne[47]. Programy *prebunkingu* wdrożono m.in. w Estonii, Finlandii, Szwecji, na Tajwanie i w Polsce (NASK-PIB, kampania „Włącz weryfikację"). Odporność społeczna nie jest jednak wyłącznie funkcją programów szkolnych. Wymaga także silnych publicznych mediów, wysokiego poziomu zaufania do instytucji i praktykowania czytania długich form.

Kluczową rolę odgrywa wykrywanie techniczne. Standard C2PA (Coalition for Content Provenance and Authenticity), rozwijany przez konsorcjum Adobe, Microsoft, BBC i innych, wprowadza kryptograficzne podpisy metadanych dla treści cyfrowych. Publikowana w formacie *Content Credentials* informacja pozwala zweryfikować pochodzenie zdjęcia lub wideo od momentu wykonania. Wdrożenie jest w toku, natomiast krytycy zwracają uwagę, że metadane są łatwo usuwalne i standard chroni przede wszystkim „prawdę pozytywną" (uwierzytelnienie autentyku), a nie „prawdę negatywną" (identyfikację falsyfikatu). Niewidzialne znaki wodne (watermarking) w treściach generowanych przez modele obrazowe (Google SynthID) i tekstowe (znaki wodne w LLM) stanowią kolejną warstwę obrony. Ich skuteczność wobec adwersarza z dostępem do modeli *open-source* jest ograniczona[48].

Nie bez znaczenia jest także współpraca sektorowa. Europejskie Obserwatorium Mediów Cyfrowych (EDMO), sieć NASK-Polska, IFCN oraz krajowe koali-

[45] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (akt o sztucznej inteligencji). Dziennik Urzędowy Unii Europejskiej, L, 2024/1689.

[46] M. G. Sessa, R. Miguel, *The Doppelganger case: Assessment*.

[47] S. Lewandowsky, S. van der Linden, *Countering misinformation and fake new*.

[48] V. Wirtschafter, *The impact of generative AI in a global election year*, Brookings Institution 2024.

cje *fact-checkingowe* stanowią infrastrukturę, która – w połączeniu z wymogami DSA – zaczyna tworzyć „system wczesnego ostrzegania".

7. Wnioski

Rozwój badań nad AI i dezinformacją, choć intensywny, ma kilka wyraźnych luk. Po pierwsze, brakuje długoterminowych badań podłużnych mierzących skutki ekspozycji na treści AI-generatywne. Większość obecnych ustaleń pochodzi z eksperymentów krótkoterminowych, których zewnętrzna trafność (*external validity*) jest ograniczona. Po drugie, komunikatory szyfrowane (WhatsApp, Telegram, Signal, Viber) pozostają czarną skrzynką. *Deepfake* Šimečki rozprzestrzenił się głównie w kanałach Telegram. Po trzecie, „dywidenda kłamcy" jako zjawisko empiryczne wciąż nie doczekała się rygorystycznych badań ilościowych. Brakuje systematycznych szacunków, jak często i w jakich kontekstach politycy skutecznie kwestionują autentyczność prawdziwych nagrań. Po czwarte, brakuje pogłębionej analizy roli polityków-amplifikatorów. De Nadal i Jančárik zwracają uwagę, że w wielu przypadkach kluczowy nie jest sam fakt istnienia *deepfake*'a, ale to, że legitymowani politycy dzielą się nim jako autentycznym materiałem, dodając mu tym samym wagi. Po piąte, badania są mocno przechylone w stronę anglosaskiego kontekstu. Mniej wiemy o krajach globalnego Południa, gdzie infrastruktura *fact-checkingowa* jest słabsza, języki mniej reprezentowane w danych uczących, a operacje wpływu potencjalnie skuteczniejsze. W końcu, mierzenie realnego wpływu operacji wpływu na wynik wyborów pozostaje metodologicznym wyzwaniem. Rzadko dostępne są dane behawioralne łączące ekspozycję na treści z decyzjami wyborczymi, a agregatowe korelacje nie ustalają związków przyczynowych.

Przegląd literatury i studia przypadków skłaniają autorkę do czterech wniosków. Teza o „nowej erze dezinformacji" powinna być formułowana ostrożnie. Sztuczna inteligencja nie stworzyła zjawiska propagandy, ale zmieniła jej ekonomikę i geografię: obniżyła koszty produkcji, zwiększyła personalizację, rozszerzyła listę potencjalnych aktorów. Zjawisko dezinformacji ma jednak długą historię, a przypadki takie jak Słowacja, Tajwan czy Argentyna dowodzą, że skuteczność operacji zależy od podłoża społecznego bardziej niż od poziomu zaawansowania technologicznego.

Nie należy jednak popadać w drugą skrajność. Skala zagrożenia rośnie w tempie, w którym instytucje demokratyczne mają trudność za nim nadążać. *Robocall* Bidena w New Hampshire (koszt jednorazowej operacji: niewiele ponad 500 USD) pokazuje, że próg wejścia dla aktora chcącego zakłócić wybory jest dziś bardzo niski.

Poważne konsekwencje polityczne mają charakter długofalowy. Kluczowym zagrożeniem nie jest pojedynczy incydent, ale kumulatywna erozja zaufania obywateli do możliwości odróżnienia prawdy od fałszu. Kiedy debata publiczna traci wspólny fundament faktów, przestaje być deliberacją i staje się serią równoległych monologów.

Przeciwdziałanie wymaga wielowymiarowej strategii. AI Act i DSA stanowią pierwsze próby regulacyjnego uporządkowania sfery, natomiast bez inwestycji w edukację medialną, silne media publiczne, transparentność platform i międzynarodową koordynację reżimy prawne pozostaną martwą literą.

Na koniec warto sformułować pytanie o kierunek. W debacie publicznej dominuje ton „obronny": jak powstrzymać AI-dezinformację? Ważniejsze może być pytanie ofensywne: jak wykorzystać te same technologie do wzmocnienia infrastruktury weryfikacji, uczestnictwa obywatelskiego i wielojęzycznego dziennikarstwa. Krajobraz technologiczny nie zawróci; polityki reaktywne pozostaną w tyle za każdą kolejną generacją modeli. Odporność demokracji budowana jest nie w reakcji na incydenty, ale w długim namyśle nad tym, co czyni ją odporną strukturalnie: silne instytucje, wysokie zaufanie, edukacja krytycznego myślenia, kultura publicznej deliberacji. Bez tych fundamentów żadna regulacja – najlepiej nawet skonstruowana – nie ochroni obywatelskiej rozmowy przed tym, co Wardle i Derakhshan nazwali „zanieczyszczeniem informacyjnym".

Bibliografia

References

Allcott H., Gentzkow M., *Social media and fake news in the 2016 election*, „Journal of Economic Perspective" 2017, t. 31, nr 2, s. 211–236.

Bail C. A., *Breaking the social media prism: How to make our platforms less polarizing*, Princeton University Press 2021.

Bessi A., Ferrara E., *Social bots distort the 2016 U.S. presidential election online discussion*,„First Monday" 2016, t. 21, nr 11, https://firstmonday.org/ojs/index.php/fm/article/view/7090 [dostęp: 12.07.2026].

Bohacek M., *Slovakia as the precursor to deepfake-enabled election interference: Lessons learned and pathways forward,* https://workshop-proceedings.icwsm.org/abstract.php?id=2024_67 [dostęp: 12.07.2026].

Bradshaw S., Howard P. N., *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*, „Working Paper" 2019, nr 3.

Buchanan B., Lohn A., Musser M., Sedova K., *Truth, lies, and automation: How language models could change disinformation*, Center for Security and Emerging Technology, Georgetown University, https://cset.georgetown.edu/publication/truth-lies-and-automation/ [dostęp: 12.07.2026].

Chesney R., Citron D. K., *Deep fakes: A looming challenge for privacy, democracy, and national security*, „California Law Review" 2017, t. 107, nr 6, s. 1753–1820.

De Nadal L., Jančárik P., *Beyond the deepfake hype: AI, democracy, and „the Slovak case„*, „Harvard Kennedy School (HKS) Misinformation Review" 2024, t. 5, nr 4, https://eprints.gla.ac.uk/333811/1/333811.pdf [dostęp: 12.07.2026].

Ferrara E., *Bots, elections, and social media: A brief overview*, s. 95–114, w: K. Shu, S. Wang, D. Lee, H. Liu (Eds.), *Disinformation, misinformation, and fake news in social media*, Springer 2020.

Hung C.-L., Fu W.-C., Liu C.-C., *AI disinformation attacks and Taiwan's responses during the 2024 presidential election*, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf [dostęp: 12.07.2026].

Kreps S. E., Kriner D. L., *Assessing misinformation recall and accuracy perceptions: Evidence from the COVID-19 pandemic*, „Harvard Kennedy School (HKS) Misinformation Review" 2023, t. 4, nr 5.

Kreps S., McCain R. M., Brundage M., *All the news that's fit to fabricate: AI-generated text as a tool of media misinformation,* „Journal of Experimental Political Science" 2020, t. 9, nr 1, s. 104–117.

Lewandowsky S., van der Linden S., *Countering misinformation and fake news through inoculation and prebunking*, „European Review of Social Psychology" 2021, t. 32, nr 2, s. 348–384.

Mroczka K., Biderma B., *Fake newsy i deepfaki w kampanii wyborczej w Polsce w 2023 roku – case studies*, „Studia Politologiczne" 2025, nr 74, s. 23–52.

Nicas J., Herrera L. C., *Is Argentina the First A.I. Election?,* „The New York Times", 14 listopada 2023.

Nyhan B., Reifler J., *When corrections fail: The persistence of political misperceptions*, „Political Behavior" 2010, t. 32, nr 2, s. 303–330.

OECD.AI., *AI-generated deepfakes manipulate Argentine presidential election*. OECD AI Incidents and Hazards Monitor, https://oecd.ai/en/incidents/2023-11-14-1e5d [dostęp: 12.07.2026].

Ognyanova K., Lazer D., Robertson R. E., Wilson C., *Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power*, „Harvard Kennedy School (HKS) Misinformation Review" 2020 t. 1, nr 4.

Paris B., Donovan J., *Deepfakes and cheap fakes: The manipulation of audio and visual evidence*, https://datasociety.net/research-library/deepfakes-and-cheap-fakes-the-manipulation-of-audio-and-visual-evidence/ [dostęp: 12.07.2026].

Pennycook G., Rand D. G., *The psychology of fake news*, „Trends in Cognitive Sciences" 2021, t. 25, nr 5, s. 388–402.

Persily N., Tucker J. A. (Eds.), *Social media and democracy: The state of the field, prospects for reform*, Cambridge 2020.

Piontek D., *Wyborcy i media. Ocena działań mediów w okresie polskiej parlamentarnej kampanii wyborczej 2023 roku*, „Przegląd Politologiczny" 2025, t. 30, nr 1.

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych (akt o usługach cyfrowych). *Dziennik Urzędowy Unii Europejskiej*, L 277/1.

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (akt o sztucznej inteligencji). *Dziennik Urzędowy Unii Europejskiej*, L, 2024/1689.

Sessa M. G., Miguel R., *The Doppelganger case: Assessment of platform regulation on the EU disinformation environment*, NATO Strategic Communications Centre of Excellence 2024.

Simon F. M., Altay S., Mercier H., *Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown*, „Harvard Kennedy School (HKS) Misinformation Review" 2023, t. 4, nr 5.

Slimovich A., *Desinformación e inteligencia artificial en las redes sociales en la Argentina*, w: D. H. Cabrera-Altieri, G. López-García & E. Campos-Domínguez (red.), *Perturbaciones informativas. Desinformación y mediatización digital*. Comunicación Social Ediciones y Publicaciones, 2025.

Summary of the Information Threat to the Paris 2024 Olympic and Paralympic Games, https://www.sgdsn.gouv.fr/files/files/Publications/20240919_NP_SGDSN_VIGINUM_Summary%20information%20threat%20Paris2024Games_EN_0.pdf [dostęp: 12.07.2026].

Świeboda H., *Generatywna sztuczna inteligencja wyzwaniem dla bezpieczeństwa społecznego i państwa,* „Maritime Security Yearbook" 2025, t. 19.

Vosoughi S., Roy D., Aral S., *The spread of true and false news online,* „Science" 2018, t. 359, nr 6380, s. 1146–1151.

Wang A. H.-E., *AI-generated disinformation in the 2024 Taiwan presidential election,* Friedrich Naumann Foundation for Freedom, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf [dostęp: 12.07.2026].

Wardle C., Derakhshan H., *Information disorder: Toward an interdisciplinary framework for research and policy making* (DGI(2017)09). Council of Europe, https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html [dostęp: 12.07.2026].

West D. M., *How disinformation defined the 2024 election narrative*, https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/ [dostęp: 12.07.2026].

Wirtschafter V., *The impact of generative AI in a global election year*, Brookings Institution 2024.