


## Człowiek i maszyna w walce z fake news: porównanie analizy cech lingwistycznych dezinformacji dokonanej przez człowieka i sztuczną inteligencję - narzędzie uczenia maszynowego

Human and machine in the fight against fake news: a comparison of analysis of linguistic features of disinformation made by a human and artificial intelligence - a machine learning tool

Aleksandra PAWLICKA

Uniwersytet Warszawski/ University of Warsaw

E-mail: a.pawlicka5@uw.edu.pl, 

**Abstract:** The term ‘fake news’ is now firmly established in public discourse and collective consciousness; Internet disinformation is a serious problem which is capable of shaking the foundations of democracy. One method of detecting fake news is to use machine learning techniques; ideally, these tools should be ‘explainable’. The aim of this paper is to present a set of linguistic features indicative of fabrication of news, to perform a human analysis of these features, to determine the veracity messages by means of artificial intelligence – a machine learning tool, and to test whether a human researcher and the machine learning algorithm recognize fake news by paying attention to the same linguistic features of the messages.

**Keywords:** disinformation, fake news, linguistic features, misinformation

### Wstęp

Rozwój Internetu przyniósł ludziom niezliczone korzyści, a także w nieodwracalny sposób zmienił to, jak pracujemy, robimy zakupy czy spędzamy wolny czas. Wraz z narodzinami Internetu pojawiły się nowe negatywne zjawiska i zagrożenia, które nie istniały przed erą Internetu albo wraz z nim nabrały nowego znaczenia. Takim zjawiskiem są między innymi *fake news*, tj. fałszywe wiadomości (nieprawdziwe informacje), które już kilka lat temu zaczęły wkraczać do publicznego dyskursu i zbiorowej świadomości. Chociaż nieprawdziwe treści są znane od czasu, kiedy pierwsi ludzie zaczęli komunikować się, dopiero rozwój Internetu stał się wyjątkowo żyznym gruntem dla ich rozpowszechniania. Kłamstwa w sieci są wyjątkowo groźne, ponieważ ich skutki wykraczają daleko poza cyberprzestrzeń – mogą rzucić na kogoś podejrzenie, zszargać jego reputację, zniszczyć karierę. Nie szkodzą one tylko pojedynczym osobom – ostatnie lata pokazały, że umiejętnie zaimplementowana dezinformacja ma moc zasiania wątpliwości w decyzje rządu, zmiany wyników demokratycznych wyborów, wywołania zamieszek. Internetowa porada, by leczyć infekcję koronawirusem poprzez wypicie wybielacza do toalet brzmi absurdalnie, a jednak łączy się z nią setki zgonów zdesperowanych ludzi, którzy w nią uwierzyli (M.S. Islam i in. 2020). Jeżeli

*fake news* podawane są obywatelom w sposób intencjonalny i zaplanowany, są w stanie zachwiać podstawami demokracji (G. Mills 2019; K. Gradoń 2020; K.A. Soomro i in. 2020).

Termin „fake news” zyskał na szczególnej popularności dzięki wyborom prezydenckim w USA w 2016 roku, których ostateczny rezultat przypisuje się celowej dezinformacji (R. Gunther/ E.C. Nisbet/P. Beck, 2018). Ich popularności pomogła także prezydentura Donalda Trumpa, który w swoich tweetach regularnie odnosił się do tego terminu, mając na myśli treści niebędące po jego myśli (D. Lind 2018). Nie ma jednak naukowego konsensusu w kwestii jednoznacznego zdefiniowania terminu „fake news”. Mówi się wręcz, że używany jest zbyt nieprecyzyjnie i w odniesieniu do zbyt różnych zakresów pojęciowych. Z tego powodu badacze i politycy postulują używanie innych określeń. Najpopularniejsze proponowane alternatywy to *disinformation* (dezinformacja, czyniona celowo) oraz *misinformation* (wprowadzanie w błąd w sposób nieświadomy) (F. Giuliani-Hoffman 2017). Pomimo tych apeli termin „fake news” w znaczeniu sfabrykowanych wiadomości na dobre zagościł w dyskursie i zbiorowej świadomości. Został on nawet wybrany słowem roku 2017 Słownika Collinsa (S. Meza 2017). Z tego powodu w niniejszej pracy będzie on używany zamiennie z dezinformacją i fabryką.

Należy zastanowić się, skąd w ludziach tak duża podatność na sieciową dezinformację. Jest ku temu kilka powodów. Część z nich związana jest z różnicami pomiędzy tradycyjnym procesem publikacji treści, np. w gazetach czy telewizji, a tym co dzieje się w Internecie. W przeciwieństwie do publikacji treści przez wydawców, zatrudniających profesjonalnych redaktorów, Internet nie ma żadnych mechanizmów kontroli treści. Po drugie, pozwala on na całkowitą anonimowość. Osoba kryjąca się pod awatarem czy pseudonimem może być ekspertem w dziedzinie albo nie wiedzieć o niej nic. Może wręcz takiego eksperta celowo udawać. Badania dowodzą, że bycie takim właśnie „leniwym czytelnikiem”, tzn. przyjmowanie treści jako pewne bez angażowania krytycznego myślenia, jest czynnikiem najsilniej warunkującym podatność na dezinformację (G. Pennycook/ D.G. Rand 2019).

W przypadku *fake news* na niekorzyść człowieka działa jeszcze kilka dodatkowych czynników, które opisuje (X. Zhou/ R. Zafarani 2018). Po pierwsze, wykorzystują naturalną skłonność ludzi do uwierzenia w treści, które są im regularnie powtarzane. Dodatkowo, algorytmy mediów społecznościowych mają tendencję do wyświetlania treści dopasowanych do poglądów i wcześniejszych wyszukiwań użytkowników, co powoduje że otrzymują oni informacje tylko w swojej „bańce informacyjnej”. To z kolei powoduje, że dostają oni głównie treści które potwierdzają ich opinie, przez co te informacje wydają się być bardziej wiarygodne. Wreszcie, ludzie wierzą w to co przeczytają, ponieważ wiadomości z założenia są takim rodzajem treści, po którym spodziewamy się autentyczności i wiarygodności (X. Zhou/ R. Zafarani 2018).

Problem *fake news* jest wciąż daleki od rozwiązania. Fałszywa wiadomość, która wkroczy do cyberprzestrzeni, jest w zasadzie nie do zatrzymania. Obecnie można mówić o dwóch filarach walki z dezinformacją: edukacji konsumentów treści i wykrywaniu dezinformacji (M. Taboada 2021). Środowisko naukowe opracowało wiele

koncepcji detekcji *fake news*. Taka detekcja polega na próbie identyfikacji sfabrykowanych treści na podstawie sprawdzania faktów lub oceny rozmaitych cech wiadomości, począwszy od jej cech lingwistycznych, stylistycznych, składniowych, frazeologicznych itd., aż po sposoby, jakimi rozchodzi się w sieci i jak długo w niej „żyje”. Można ją przeprowadzać w sposób manualny lub automatyczny. „Ręczne sprawdzanie” prawdziwości treści, przeprowadzane przez ekspertów, jest procesem dającym najbardziej wiarygodne i pewne wyniki. Jest to jednak proces niezwykle żmudny i czasochłonny. Biorąc pod uwagę ogromne ilości danych w sieci oraz prędkość, z jaką dezinformacja rozprzestrzenia się w niej, prawdopodobnie niemożliwe jest sprawdzenie wszystkich treści w taki sposób.

Dodatkowo, człowiek w obliczu dezinformacji nie może liczyć tylko na swoją intuicję – dowiedziono, że ludzka zdolność wykrywania sfabrykowanych treści, zwłaszcza w obliczu bycia otoczonym przez dezinformację, jest tylko nieco lepsza niż przypadek. Średnia dokładność mierzona w ponad 100 eksperymentach na ponad tysiącu ludzi wyniosła około 54% (X. Zhou/ R. Zafarani 2018).

Alternatywą jest zautomatyzowanie procesu detekcji przez zastosowanie algorytmów uczenia maszynowego (ML – *Machine Learning*) – prężnie rozwijającego się obszaru sztucznej inteligencji (AI – *Artificial Intelligence*). Klasyfikacja tekstu opiera się tu na NLP (*Natural Language Processing*) – przetwarzaniu języka naturalnego. Algorytmy trenowane są na dużych zbiorach wiadomości, którym nadano odpowiednie etykiety. Po treningu narzędzie otrzymuje nowe, nieznane mu dane i bazując na dotychczasowym doświadczeniu klasyfikuje wiadomość jako posiadającą cechy wskazujące na to, że jest to *fake news*. Głównymi zaletami zautomatyzowanego wykrywania dezinformacji są jego duża szybkość oraz ogromna ilość danych, które jest w stanie sprawdzić. Utrudnieniem natomiast może być tzw. problem czarnej skrzynki (*black-box problem*), tj. brak transparentności sposobu wyciągania wniosków przez algorytm oraz zbytne, wręcz ślepe poleganie człowieka na decyzjach podejmowanych przez maszynę. Rozwiązaniem dla pierwszego problemu jest konstruowanie tzw. wyjaśnialnych systemów AI (*explainable AI, xAI*). W przypadku drugiego dylematu najlepszym wyjściem jest przyjęcia podejścia hybrydowego – ludzkiej weryfikacji wyników pracy algorytmu.

Niniejszy artykuł podejmuje zagadnienie wykrywania *fake news* z dwóch perspektyw: analizy cech lingwistycznych, które wskazują na celową fabrykację treści oraz zastosowania metod uczenia maszynowego. Wyniki dwóch eksperymentów zostaną porównane w celu sprawdzenia, czy ludzka i sztuczna inteligencja próbują identyfikować *fake news* w taki sam, czy różny sposób. Wartością tej pracy jest zebranie bogatej listy cech lingwistycznych dezinformacji z najnowszej literatury przedmiotu oraz pierwsze udokumentowane podejście do porównywania sposobów w jaki dezinformację rozpoznają ludzie, a w jaki sztuczna inteligencja.

## 1. Cechy lingwistyczne sfabrykowanych treści

Teksty celowej dezinformacji i autentycznych wiadomości różnią się na wielu płaszczyznach. Pod względem językowym są to różnice dostrzegalne i powtarzalne na tyle, że można mówić o pewnych cechach lingwistycznych, które wskazują na styl celowej

dezinformacji. Jednakże fakt obecności lub braku danego elementu w nagłówku czy treści wiadomości nie wystarczy, by od razu zakwalifikować ją jako prawdziwą bądź fałszywą. Takie elementy powinny raczej wzbudzać czujność czytelnika, zwłaszcza gdy występuje ich nagromadzenie.

Długość artykułów zawierających fałszywe wiadomości jest często mniejsza, niż w wypadku treści prawdziwych (M. Taboada 2021), często mają też one krótsze zdania (M. Mahyoob i in. 2020) i występują w nich krótsze, prostsze wyrazy (D. Shariatmadari 2019; B.D. Horne/ S. Adali 2017), a nawet błędy gramatyczne lub edytorskie (M. Mahyoob i in. 2020). Cechy te wiążą się z prostszym, łatwiejszym do zrozumienia językiem stosowanym w dezinformacji, wymagającym nieco niższego poziomu wykształcenia odbiorcy, niż treści prawdziwe o analogicznej tematyce (M. Mahyoob i in. 2020; B.D. Horne/ S. Adali 2017). Prawdziwe wiadomości są zwykle bardziej poprawne pod względem interpunkcji i gramatyki oraz ogólnego dopracowania (M. Mahyoob i in. 2020; M. Taboada 2021).

Odwrotnie do treści artykułów, nagłówki artykułów pisanych jako celowa dezinformacja są często dłuższe i bardziej „napakowane” informacjami (B.D. Horne/ S. Adali 2017). Stosowany jest w nich bardziej sensacyjny, zwracający uwagę język, nazywany „klikbajtowym” (Obserwatorium Językowe Uniwersytetu Warszawskiego 2016) (ang. *clickbait* = click + bait, kliknięcie + przynęta), charakteryzujący się m.in. użyciem przysłówków nacechowanych emocjonalnie (M. Mahyoob i in. 2020). Częściej też w fake newsach znaleźć można wykrzykniki oraz wielkie litery – zarówno na początku wyrazów, jak i w formie całych wyrazów nimi pisanych (B.D. Horne/ S. Adali 2017).

Styl artykułów sklasyfikowanych jako *fake news* nierzadko jest bardziej potoczny niż w prawdziwych wiadomościach (M. Taboada 2021). Bardziej przypomina on dyskurs narracyjny (E. Francis 2013). Jednocześnie język jest zdecydowanie mniej neutralny, częściej pojawiają się w nim wyrazy o charakterze wartościującym (B.D. Horne/ S. Adali 2017), określenia wskazujące na uprzedzenia (S. Volkova i in. 2017) oraz słowa dodające emfazy (D. Shariatmadari 2019). Ton artykułów jest „zauważalnie bardziej negatywny” (J. Paschen 2019; F.T. Asr 2019) „nadmiernie emocjonalny” (D. Shariatmadari 2019), sensacyjny i nastawiony na przyciąganie uwagi (J. Tompkins 2019). Emocje, które wzbudzają *fake news* to częściej strach, obrzydzenie, zaskoczenie (S. Vosoughi i in. 2018). Słownictwo, częściej niż w prawdziwych wiadomościach, nawiązuje do niepokoju, śmierci i seksu (F.T. Asr 2019). W odróżnieniu od dezinformacji, prawdziwe wiadomości częściej wywołują uczucia oczekiwania, smutku, radości i zaufania (S. Vosoughi i in. 2018) i odnoszą się do tematyki pracy zarobkowej, pieniędzy, gospodarki, biznesu (D. Shariatmadari 2019). Dowiedziono również, że *fake news* zdecydowanie częściej sięgają po superlatywy, tj. stopień najwyższy przysłówków i przymiotników, i przesadę, podczas gdy prawdziwe wiadomości raczej uciekają się do stosowania stopnia wyższego i porównań (M. Mahyoob i in. 2020).

Pod względem gramatycznym, zdania w nieprawdziwych wiadomościach częściej zawierają konstrukcje mowy zależnej (M. Mahyoob i in. 2020), znacząco więcej

nazw własnych oraz konstrukcji czasownikowych (B.D. Horne/ S. Adali 2017) i przyśłówków (F.T. Asr 2019), a także zaimków trzecioosobowych w miejsce imion i nazwisk (D. Shariatmadari 2019). Częstsze użycie zaimków trzeciej osoby liczby mnogiej wiąże się także z tym, że dezinformacja częściej stosuje retorykę „my kontra oni”, alienując konkretne grupy i przedstawiając je w złym świetle.

Co ciekawe, odnotowano, że w fałszywych wiadomościach częściej niż w prawdziwych występuje słownictwo odnoszące się do zmysłu wzroku. Z kolei w tych prawdziwych – słownictwo odnoszące się do słuchu. Tłumaczy się to nawiązywaniem do różnicy między usłyszeniem informacji u źródła, a pamiętaniem, jak się ją fabrykowało, tj. wizualnego wspomnienia pisania o niej (M. Taboada 2021).

## 2. Opis eksperymentu

Celem badania było porównanie, czy analizy cech lingwistycznych nagłówków fake newsów przeprowadzona przez człowieka oraz przez istniejące, sprawdzone narzędzie uczenia maszynowego da takie same, czy różne wyniki.

W trakcie projektowania eksperymentu wykorzystano najnowsze doniesienia literatury dotyczące wspomnianych cech lingwistycznych, natomiast do przeprowadzenia eksperymentu wykorzystano dane ze zbioru danych opisanego w (H. Ahmed/ I. Traore/ S. Saad 2017, 2018). Jest to zbiór, którego jedna część zawiera wiadomości uznane na podstawie analizy faktów za fake newsy, a druga część – wiadomości uznane za prawdziwe. Każda z części zawiera ponad dwanaście tysięcy artykułów, zebranych głównie w latach 2016 i 2017. W większości są to wiadomości ze świata i o tematyce politycznej.

Narzędziem do generowania losowych liczb (random.org) wybrano losowe nagłówki ze zbioru danych zawierających fałszywe wiadomości. Następnie, nagłówki poddano analizie innowacyjnym narzędziem opisanym w pracy R. Kozik i in. (2022). Wykorzystano w nim głęboką konwolucyjną sieć neuronową połączoną z klasycznymi, płytkami modelami. W toku badań, jego dokładność wyrażona metryką *accuracy* (suma wyników prawdziwie dodatnich i prawdziwie ujemnych podzielona przez sumę wyników prawdziwie i fałszywie dodatnich oraz prawdziwie i fałszywie ujemnych) wyniosła 96–97%. Wybrano pierwszych 20 nagłówków, które narzędzie wskazało jako zdecydowanie wykazujące cechy wskazujące na fabrykację; wyniki niepewne lub wskazujące na prawdziwość nagłówków zostały odrzucone.

W modelu wykorzystano techniki wyjaśnialności pochodzące z teorii gier, zdolne wskazać wpływ poszczególnych cech na decyzję narzędzia. Wyniki klasyfikacji wybranych wcześniej dwudziestu nagłówków poddano analizie pod kątem wyjaśnialności, tzn. sprawdzono, które wyrazy zdaniem algorytmów wskazują na możliwą fabrykację treści.

W ostatniej części eksperymentu zestawiono ze sobą wyniki analizy nagłówków dokonanej przez narzędzia uczenia maszynowego oraz analizy cech lingwistycznych tych nagłówków, zgodnej z najnowszym stanem wiedzy, przeprowadzonej przez Autorkę. Celem było sprawdzenie, czy wyniki obu analiz pokryją się, czy też będą podkreślać inne cechy wskazujące na fabrykację treści.

### 3. Porównanie wyników analiz

W tej części artykułu ukazano zestawienie wyników eksperymentu. Tabela 1 przedstawia nagłówki wybrane do badania, wskazuje jakie słowa według narzędzia uczenia maszynowego wskazują na możliwą fabrykację treści oraz podkreśla cechy językowe bardziej typowe dla sfabrykowanych niż dla prawdziwych treści. W przypadku słów wskazanych przez narzędzie zaprezentowano wszystkie uznane za świadczące o fabrykacji. Cechy językowe podkreślone przez autorkę to te zgodne z najbardziej aktualnym stanem wiedzy zawartym w literaturze przedmiotu (M. Taboada 2021; S. Vossoughi i in. 2018).

L.p.	Treść nagłówka	Słowa wskazane przez AI	Cechy językowe wskazujące na fabrykację
1.	Comment on The Fix is In — Russia Kicked Off UN Human Rights Council, While Terrorist Saudi Arabia Re-Elected	Russia, Arabia, UN, Saudi, elected, human, terrorist	nasylenie wyrażeniami czasownikowymi, potoczny i mocny język, wywołuje poczucie zagrożenia i oburzenie
2.	EVIL HILLARY SUPPORTERS Yell "F*ck Trump"...Burn Truck Of Daddy Fishing With 2 Yr Son Over Of Trump Bumper-Stickers [VIDEO]	Hillary, evil, supporters, Trump	przekleństwa, sensacyjny język, nieformalny skrót, słowa o charakterze wartościującym, język potoczny, wywołuje oburzenie
3.	Re: At This Current Pace, A Record-Shattering 2.4 Trillion Dollars Will Be Added To The National Debt This Year	re, this, pace, dollars, debt	słownictwo ekstremalne, nawiązanie do zagrożenia, wielkie litery na początkach wszystkich wyrazów
4.	Treason! NYT vows 'rededication' to reporting!	reporting	wykrzykniki, sensacyjny wydźwięk, wywołuje oburzenie
5.	Progressives Find 'White Trash' More Threatening Than Nuclear War	threatening, progressives, war, nuclear	określenia wulgarne i potoczne, wielkie litery na początkach wszystkich wyrazów, słownictwo sensacyjne, nawiązujące do grozy, niepokoju
6.	Dozens Of Syrian Children Dead After Receiving Measles Vaccinations	dozens, children, Syrian	sensacyjny język, nagłówek wywołujący poczucie grozy i oburzenia, nawiązuje do śmierci
7.	Purely Coincidental or the Path to a Stolen Election? Voters Report 'Irregularities' with Voting Machines Across the US	voting, stolen, machines, election, path, report, across, voters	mocne słownictwo, implikacje wywołujące oburzenie, wielkie litery na początkach wszystkich wyrazów
8.	Comment on Hollyweird Clinton supporter was so poor she couldn't afford to buy designer shoes by Dr. Eowyn	Clinton, supporter, designer,	prześmiewcze imię, nagromadzenie nazw własnych, sensacyjny, klikbajtowy wydźwięk

		couldn't, afford, dr, poor	
9.	Four Common Mistakes When Burning Wood	burning	klikbajtowy wydźwięk nagłówka, wielkie litery na początkach wszystkich wyrazów
10.	Fundamentally transforming the, uh, um, whatever...	whatever	potoczny, satyryczny wydźwięk, nieformalne słownictwo
11	BREAKING : Mike Pence DEMANDS the FBI RELEASE ALL EMAILS BEFORE ELECTION DAY	to, before, mike, emails, breaking, election, all	sensacyjny i mocny wydźwięk, wyrazy zapisane wielkimi literami, usterki edycyjne
12.	RIGGED! DOJ Weiner Email Probe Will Be Led By PODESTA'S CLOSE FRIEND Peter Kadzik	rigged, email, doj, weiner, probe, close	sensacyjny i mocny wydźwięk, wyrazy zapisane wielkimi literami, nagromadzenie imion, nazwisk i nazw własnych, wykrzyknik, wywołuje uczucie oburzenia
13.	Bill Clinton Wants to be Called Something Completely Ridiculous if Hillary is Elected	Hillary, Clinton, completely, Bill, wants	klikbajtowy wydźwięk nagłówka, wielkie litery na początkach wszystkich wyrazów, wartościujące określenia, słowa emfaticzne, natężenie wyrażen czasownikowych
14.	What If I Told You Cannabis Is Great for Reversing Alzheimer's?	what, cannabis, you	sensacyjny, klikbajtowy wydźwięk, wielkie litery na początkach wszystkich wyrazów, zwrócenie się do czytelnika w drugiej osobie liczby pojedynczej, określenia wartościujące
15.	Fake Cigarettes are Being Sold and Killing People, Here's how to Spot Counterfeit Packs	being	sensacyjny, klikbajtowy wydźwięk, nawiązuje do śmierci i zagrożenia, wielkie litery na początkach wszystkich wyrazów
16.	Confusing Jihad with Hirabah Won't Build a More Peaceful World	peaceful, world	nawiązanie do zagrożenia, wielkie litery na początkach wszystkich wyrazów
17.	Obama Legacy: Medical Tyranny? New Order Issued for Detainment and Quarantines	Obama, order, legacy, medical	wywołuje oburzenie, słownictwo ekstremalne i wartościujące, wszystkie wyrazy napisane wielkimi literami, sensacyjny wydźwięk

18.	Are Microwave Transmission Weapons Of Mass Destruction Being Used To Trigger Catastrophic Earthquakes?	destruction, weapons, used, being, mass	słownictwo ekstremalne, emfatyczne, sensacyjny wydzźwięk wywołujący grozę, nawiązanie do śmierci i zagłady
19.	Is Feminism A Giant Shit Test?	shit	wulgarnie słownictwo potoczne, wielkie litery na początkach wszystkich wyrazów, słownictwo ekstremalne
20.	Dad Throws His Son Off Bridge for Breaking His Apple Watch	son, throws, off, bridge, apple, breaking, his	określenia potoczne, nazwa własna, nawiązanie do śmierci i grozy, wszystkie wyrazy napisane wielkimi literami

Tabela 1. Zestawienie cech lingwistycznych wskazujących na dezinformację oraz słów wskazanych przez algorytm sztucznej inteligencji.

Na podstawie zgromadzonych danych można stwierdzić, że cechy nagłówków, które wpływają na przekonanie o ich sfabrykowaniu, są różne dla narzędzia uczenia maszynowego i dla badacza. Ludzki badacz, analizując cechy lingwistyczne dezinformacji, skupia się na słowach, które wywołują konkretne emocje, które nie pasują do stylu, nie są neutralne. Algorytm sztucznej inteligencji o możliwej dezinformacji decyduje, gdy nagłówek zawiera wyrazy występujące wcześniej w innych wiadomościach, sklasyfikowanych jako *fake news* – głównie nazwiska i nazwy własne, w połączeniu z konkretnymi czasownikami, rzeczownikami, przymiotnikami i przysłówkami.

#### 4. Wnioski

Na podstawie wyników eksperymentu uwidoczniło się, że analizując dezinformację pod względem lingwistycznym, człowiek i algorytmy sztucznej inteligencji zwracają uwagę na inne elementy. Człowiek, oceniając prawdopodobieństwo fabrykacji treści, zwraca uwagę na cechy edytorskie nagłówka, konkretne części mowy w nim zawarte oraz emocje, które nagłówek stara się wzbudzić w czytelniku. Narzędzie uczenia maszynowego wybiera konkretne słowa, które występowały w tytułach innych *fake news*ów.

Z wyników badań można wnioskować, że o ujawnieniu dezinformacji mogą decydować subtelniejsze, mniej oczywiste wzory i korelacje, niż te do tej pory opisywane w literaturze przedmiotu. Jest to inspiracja do dalszych prac.

Kolejnymi etapami badań nad tym zagadnieniem będą: przetestowanie większej liczby wiadomości, innych zbiorów danych, badania nad analizą sentymentu oraz zastosowanie innych algorytmów wykrywających *fake news*, a także innych mechanizmów wyjaśnialności.

#### Bibliografia

Ahmed, H./ I. Traore/ S. Saad (2017), *Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques*, (w:) I. Traore/ I. Woungang/



- A. Awad (red.), *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*. ISDDC 2017. Lecture Notes in Computer Science, vol. 10618. Springer, Cham.
- Ahmed, H./ I. Traore/ S. Saad (2018), *Detecting opinion spams and fake news using text classification*, (w:) „Security and Privacy” 1(1), e9.
- Asr, F.T. (2019), *The language gives it away: How an algorithm can help us detect fake news*. The Conversation.
- Francis, E. (2013), *MisInfoWars: A Linguistic Analysis of Deceptive and Credible News*. Praca magisterska.
- Giuliani-Hoffman, F. (2017)., *“F\*\*\* News” should be replaced by these words, Claire Wardle says*, (w:) „CNN Business,, (URL <https://money.cnn.com/2017/11/03/media/claire-wardle-fake-news-reliable-sources-podcast/index.html>). [Pobrano: 22.02.2022].
- Gradoń, K. (2020), *Crime in the time of the plague: fake news pandemic and the challenges to law-enforcement and intelligence community*, (w:) „Society Register” 4(2), 133–148.
- Gunther, R/ E.C. Nisbet/ P. Beck (2018), *Trump may owe his 2016 victory to ‘fake news,’ new study suggests*. The Conversation.
- Horne, B.D./ S. Adali (2017). *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News*. (URL <http://arxiv.org/abs/1703.09398>). [Pobrano: 22.02.2022].
- Islam, M.S. i in. (2020), *COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis*, (w:) „The American Journal of Tropical Medicine and Hygiene” 103(4), 1621–1629.
- Kozik, R. i in. (2022), *Fake News Detection Platform – Conceptual Architecture and Prototype*, (w:) „Logic Journal of the IGPL” 30(6), 1005–1016.
- Lind, D. (2018), *President Donald Trump finally admits that “fake news” just means news he doesn’t like*, (w:) „Vox” (URL <https://www.vox.com/policy-and-politics/2018/5/9/17335306/trump-tweet-twitter-latest-fake-news-credentials>). [Pobrano: 22.02.2022].
- Mahyoob, M./ J. Algaraady/ M. Alrahaili (2020), *Linguistic-Based Detection of Fake News in Social Media*, (w:) „International Journal of English Linguistics” 11(1), 99–109.
- Meza, S. (2017), *“Fake News” Named Word of the Year*, (w:) „Newsweek” (URL <https://www.newsweek.com/fake-news-word-year-collins-dictionary-699740>). [Pobrano: 22.02.2022].
- Mills, G. (2019), *Why LE needs to prepare for the disinformation era*, (w:) „Police 1” (URL <https://www.police1.com/chiefs-sheriffs/articles/why-le-needs-to-prepare-for-the-disinformation-era-8AgmShfpEXwW1RrK/>). [Pobrano: 22.02.2022].
- Obserwatorium Językowe Uniwersytetu Warszawskiego (2016), *klikbajtowy*, (URL <https://nowewyrazy.uw.edu.pl/haslo/klikbajtowy.html>). [Pobrano: 22.02.2022].
- Paschen, J. (2019), *Investigating the emotional appeal of fake news using artificial intelligence and human contributions*, (w:) „Journal of Product & Brand Management” 29(2), 223–233.

- Pennycook, G. / D.G. Rand (2019), *Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning*, (w:) „Cognition” 188, 39–50.
- Shariatmadari, D. (2019), *Could language be the key to detecting fake news?*, (w:) „The Guardian” (URL <https://www.theguardian.com/commentisfree/2019/sep/02/language-fake-news-linguistic-research>). [Pobrano: 22.02.2022].
- Soomro, K.A. i in. (2020), *Digital divide among higher education faculty*, (w:) „International Journal of Educational Technology in Higher Education” 17(1), 21.
- Taboada, M. (2021), *Authentic Language in Fake News*, (w:) „Items” (URL <https://doi.org/https://items.ssrc.org/beyond-disinformation/authentic-language-in-fake-news/>). [Pobrano: 22.02.2022].
- Tompkins, J. (2019), *Disinformation Detection: A review of linguistic feature selection and classification models in news veracity assessments*. (URL <http://arxiv.org/abs/1910.12073>). [Pobrano: 22.02.2022].
- Volkova, S. i in. (2017), *Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter*, (w:) „Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics” volume 2: Short Papers, 647–653.
- Vosoughi, S./ D. Roy/ S. Aral (2018), *The spread of true and false news online*, (w:) „Science” 359(6380), 1146–1151.
- Zhou, X./ R. Zafarani (2018), *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*, (w:) „ACM Comput. Surv.” 53(5).