

Bartłomiej Balcerzak

NAJLEPSZE NARZĘDZIE DO ZWALCZANIA FEJKÓW (I DLACZEGO NIE DZIAŁA)

STRESZCZENIE

Artykuł stanowi przybliżenie technik informatycznych wykorzystywanych w procesie rozpoznawania fałszywych wiadomości. Równocześnie wskazuje na społeczne i psychologiczne czynniki wpływające na popularność fake news, oraz wskazuje na społeczne konsekwencje stosowania różnych metod wykrywania Fake News.

Słowa kluczowe: fake news, informatyka społeczna, przetwarzanie języka naturalnego, uczenie maszynowe

WPROWADZENIE

W roku 2020 pojęcie *fake news* jest już dobrze zdomowione w słowniku każdego, kto nawet w najmniejszym stopniu jest zainteresowany tym, jak przebiega debata publiczna, zarówno na poziomie poszczególnych państw, jak i w skali globalnej. Oczywiście, manipulacje, fałszerstwa lub najzwyczajniejsze kłamstwa stanowią element dyskusji od tysiącleci (do tego nawet stopnia, że złośliwy erudyta mógłby wypełnić kilka stron manuskryptu samymi tylko sławnymi przykładami oszustw ze wszystkich epok), jednakże w dobie Internetu problem fałszywek nabrał nowego znaczenia. Po pierwsze, nastąpiła zmiana ilościowa, treści medialne produkowane są w stopniu nieznanym wcześniejszym mediom, dotyczą wszystkich aspektów życia. Nic więc dziwnego, że w takim liczyonym w Petabajtach

danych strumieniu informacji fałszywki – nawet jako ułamek produkowanych treści – będą stanowiły zbiór manipulacji i kłamstw.

O wiele poważniejszą zmianą jest różnica jakościowa, która objawia się na wielu płaszczyznach. Po pierwsze zmianie uległ sposób w jaki informacja jest produkowana. W czasach przedinternetowych produkcja treści wymagała środków, infrastruktury oraz nakładów czasu i pracy. Gazety, książki, telewizja, media starożytności są scentralizowane, mają swoich wydawców, dziennikarzy, informatorów. W dodatku jeżeli informacja była fałszywa, można było wskazać źródło przekłamania. Nawet w społeczeństwach totalitarnych możliwe było rozpoznanie, czy w wiadomościach nie ma prawdy, a w prawdzie nie ma wiadomości¹. W dobie sieci 2.0 treści tworzyć może każdy (niezależnie od tego, jaką jakość ta treść prezentuje). Co więcej, twórca treści nie musi być nawet człowiekiem, może być botem, który odtwarza zaprogramowane treści². Prowadzi to do sytuacji, w której informacja rozprzestrzenia się w sposób zdecentralizowany, bez wskazanego źródła. Wydawać by się mogło, że w takim środowisku informacja może rozprzestrzeniać się łatwiej, a wszelkie kłamstwa można łatwo weryfikować. Rzeczywistość jednak wskazuje coś zupełnie innego. Łatwo wymienić tutaj sytuacje, w których informacje fałszywe (bądź manipulowane) z łatwością rozprzestrzeniają się w Internecie. Z całego morza teorii spiskowych, łańcuszków maili, szokujących linków wystarczy wspomnieć kilka przykładów z ostatnich miesięcy.

10

W marcu tego roku, kiedy w Polsce pojawiały się pierwsze przypadki COVID-19, na Messengerze i w innych komunikatorach internetowych zawrzało od „potwierdzonych” informacji od krewnego, kolegi, sąsiadki, zapewniających, że rząd zamierza zamknąć miasta, wprowadzić stan wyjątkowy, wyprowadzać na ulice wojsko. Nie trzeba było długo czekać, by do łańcuszków dołączyły „sprawdzone” informacje o cudownych lekarstwach na koronawirusa albo zapewnienia, że wirusa tak naprawdę nie ma i wszystko to jeden wielki spisek. Spisek, powiadacie? Oczywiście, przecież widzieliśmy wszyscy linki mówiące o tym, że obecna pandemia to efekt stawiania masztów 5G, która to technologia miała służyć do kontroli umysłów, redukcji populacji zgodnie z wolą Billa Gatesa (i jego morderczych szczeponiek).

Teorie spiskowe, które zresztą same w sobie są tematem fascynującym i o wiele głębszym niż tylko foliowe kapelusze i płaska ziemia, dzięki Internetowi przeżywają renesans. Od wielogodzinnych filmów o spiskach Iluminatów (obrazowo

nazywanych w środowiskach polskich fanów tego medium jako filmy „z żółtymi napisami”) po dzisiejsze teorie propagowane przez memy zanurzone w wielu warstwach ironii, teorie spiskowe stanowią trwały element internetowej debaty publicznej.

Jednakże esej swój zacząłem od pojęcia *fake news*, czemu zatem skupiam się teraz na teoriach spiskowych? Nie zamierzam tutaj stawiać znaku równości między tymi dwoma pojęciami. Wszak jedno z nich oznacza określony

typ produkowanej w Internecie treści, takiej, która zgodnie z najpopularniejszymi definicjami jest intencjonalnie fałszywa, wzbudzająca emocje, a zarazem prawdopodobna i trudna do weryfikacji³. Za to drugie pojęcie stanowi pewien całościowy obraz rzeczywistości, w którym kluczowe wydarzenia są rezultatem działalności bytów, których istnienie oraz sposób funkcjonowania są niejawne, za to motywacje są jawnie złe⁴.

Nie zamierzam tu również postulować, że każda teoria spiskowa składa się z samych *fake newsów*. Byłoby skrajną naiwnością twierdzić, że wpływowi ludzie nigdy nie dogadywali się w sekrecie, by osiągnąć swoje cele kosztem innych. To też nie jest tu moim celem.

Chciałbym jedynie wskazać, że obydwie zjawiska łączą się ze sobą i są rezultatem zarówno właściwości używanej przez nas technologii, jak i efektów psychologicznych na które wszyscy jesteśmy podatni. Czy tego chcemy, czy nie, w Internecie od fejków i teorii spiskowych nie uciekniemy. Nie oznacza to jednak, że nie było i nie ma prób aby ich propagację ograniczyć.

Czy tego chcemy, czy nie, w Internecie od fejków i teorii spiskowych nie uciekniemy. Nie oznacza to jednak, że nie było i nie ma prób aby ich propagację ograniczyć.

INFORMATYKA W WALCE Z FAKE NEWS - FACT-CHECKING

Oczywiście, pierwszym pomysłem jaki pojawia się, kiedy rozważane są narzędzia zwalczające *fake newsy*, byłoby rozwiązanie najprostsze, czyli sprawdzanie faktów (*fact-checking*). Pomysł jest prosty: tworzymy platformę na której można zgłaszać różne treści, materiały i spotykane w Internecie stwierdzenia. Następnie działający na platformie specjaliści oceniają, czy materiały te są wiarygodne, czy nie.

Obecnie istnieje już kilka takich usług, w kontekście polskim taką platformą jest Demagog⁵ poświęcony przede wszystkim weryfikacji wypowiedzi polityków. Jest on odpowiednikiem amerykańskiej platformy Politifact⁶, która w analogiczny sposób ocenia wypowiedzi polityków w USA. Innym przykładem systemu oceny faktów, zresztą jednym z najstarszych, jest strona Snopes.com, która zawiera oceny prawdziwości teorii spiskowych, legend miejskich oraz innych treści znanych z Internetu bądź innych mediów. Z pozoru takie rozwiązanie brzmi całkowicie sensownie, jednakże wobec *fake newsów* jest niewystarczające. Z kilku powodów.

Pierwszym problemem jest kwestia jakości pracy recenzentów, jak i w ogóle zapewnienie właściwych recenzentów do zadania. W jednych dziedzinach, takich jak medycyna, zadanie to jest relatywnie łatwe. W innych zadanie takie może być jednak o wiele trudniejsze. Dobrym przykładem *fake newsów*, przy których weryfikacja byłaby trudna, mogą być informacje dotyczące jednostkowych zdarzeń. Taką informacją mogą być np. doniesienia o pobiciu dokonany przez grupę uchodźców; informacja o tym, że działacz określonej frakcji politycznej stał się ofiarą ataku albo modne w pierwszych dniach pandemii łańcuszki o zamknięciu miast przez rządy. Nie mamy tutaj eksperta, który rozsądzi, czy opisywane zdarzenie definitywnie miało miejsce. Zresztą, nawet jeśli kłamstwo zostanie wskazane, cel *fake newsa* został osiągnięty – wzmocnił określoną narrację na konkretny temat. A narracje mogą żyć poza faktami.

12

Innym dobrym przykładem takiej trudności może być analiza przypadku niekoniecznie powiązanego z problemem *fake news* (ale stanowiącego przejaw tego samego szerszego procesu cywilizacyjnego), czyli popularności alternatywnych teorii dziejów, której chyba najbardziej reprezentatywną formą jest teoria starożytnych kosmitów⁷. Popularność tej teorii bierze się nie tylko z faktu, że jest „fajna” (co zresztą jest często pojawiającą się cechą internetowych fejków – są „fajne”, ale o tym jeszcze później), ale również z faktu, że weryfikacja twierdzeń jej zwolenników wymaga specjalistycznej wiedzy. Jeżeli nie znamy języków starożytnego świata lub nie śledzimy stanu wiedzy archeologicznej, nie będziemy w stanie stwierdzić, czy Sumerowie naprawdę nazywali swój kraj „Krainą Królów o płonących raketach” i założyli religię Zoroastryzmu. Na świecie nie ma zbyt wielu sumerologów, wielu z nich też nie miałoby czasu bądź chęci, by zajmować się ciągłym wyjaśnianiem spraw, które odciągają ich od właściwych badań.

Jak widać, kwestia dostępności właściwych ekspertów może stanowić wyzwanie, ale innym, jeszcze poważniejszym problemem dotyczącym ekspertów w *fact-checkingu* może być właściwe uzasadnienie ich kompetencji. O ile w pewnych dziedzinach łatwo o obiektywność (nauki ścisłe, logika), o tyle w sprawach najbardziej „fejkogennych”, takich jak polityka, ideologia bądź religia trudno o bezstronną ewaluację, gdyż trudno też o bezstronność. Nawet najbardziej życzliwie nastawiony recenzent fejków dotyczących uchodźców, może być podważany jako propagator określonej ideologii bądź politycznej frakcji.

Inną kwestię, która sprawia, że platformy *fact-checkingowe* są niewystarczające w walce z fejkami, można zdefiniować słowami: „reaktywność” i „szybkość”. A właściwie to brakiem tych przymiotów. Bowiem z samej natury faktów wynika, że możemy je sprawdzać dopiero „po fakcie”. Oznacza to, że fejk już się rozposzechnił i wyrządził szkody, zanim dotarł do recenzentów. Równocześnie, mając przeciw sobie farmy botów, armie opłacanych trolli i automatyczne generatory treści, jakakolwiek platforma opierająca się na pracy ręcznej będzie na straconej pozycji.

WSPOMAGANY FACT-CHECKING

13

Pewną formą poradzenia sobie z ograniczeniami przed jakimi stoją platformy *fact-checkingowe* jest wspomaganie ich za pomocą narzędzi automatycznych. Jednym z kierunków, jaki można podjąć, jest automatyzacja procesu oceny treści poprzez filtrowanie nowych tekstów i porównywanie ich z tekstami już ocenionymi przez ekspertów. Intuicja stojąca za tym rozwiązaniem jest prosta. Zdania o zbliżonej semantyce, takie jak:

- a) Szczepionki powodują raka
i
- b) Szczepionki powodują nowotwory, powinny posiadać tę samą wiarygodność.

Brzmi łatwo, jednakże wprowadzenie takiego rozwiązania w życie to już zupełnie inna sprawa.

Pewną formą poradzenia sobie z ograniczeniami przed jakimi stoją platformy *fact-checkingowe* jest wspomaganie ich za pomocą narzędzi automatycznych. Jednym z kierunków, jaki można podjąć, jest automatyzacja procesu oceny treści poprzez filtrowanie nowych tekstów i porównywanie ich z tekstami już ocenionymi przez ekspertów.

W pierwszej kolejności musielibyśmy mieć jakąś przyjętą miarę określającą podobieństwo semantyczne oraz funkcję, która pozwoliłaby na wyliczenie takiego podobieństwa dla dowolnej pary tekstów. Pierwsza taka poważna próba została podjęta w 2012 roku w ramach corocznego wyzwania na konferencji informatycznej SemEval⁸. Zdefiniowano tam skalę podobieństwa semantycznego i przetestowano serię algorytmów – od metod opartych na częstotliwości występowania słów kluczowych, po metody oparte na ręcznie przygotowanych sieciach relacji semantycznych. Obecnie jednak głównym zestawem narzędzi wykorzystywanych w tym (i wielu innych) wyzwaniach są tzw. modele wektorowe. Są to algorytmy oparte o struktury sieci neuronowych, które uczą się znaczenia słów w oparciu o to, w jakim otoczeniu się pojawiają. Rezultatem końcowym działania takiego algorytmu jest reprezentacja każdego słowa za pomocą wektora, czyli położenia w wielowymiarowej przestrzeni. W chwili obecnej głównymi algorytmami wykorzystywanymi przez zespoły badawcze i firmy na świecie są algorytmy BERT, ELMO i RoBERTa. Zwłaszcza ten ostatni jest przydatny w szacowaniu podobieństwa semantycznego, gdyż uczy się kontekstu na poziomie zdania zamiast pojedynczych słów. W ten sposób korzystając z prostych miar odległości można łatwo policzyć semantyczne podobieństwo.

14 Jednakże nie oznacza to, że problem fejków został już rozwiązany. Po pierwsze, taka faktycznie działająca funkcja podobieństwa zachowująca wiarygodność jest dopiero w powijakach. Relacja łącząca zawartość semantyczną oraz wiarygodność danej treści nie jest jeszcze w pełni opisana. Główną przeszkodą na drodze do konstrukcji takiej funkcji jest właściwe zrozumienie pojęcia wiarygodności. Badania (co prawda ograniczone tylko do wybranych tematów medycznych) w ramach nieformalnego zespołu, którego jestem członkiem, prowadzonego pod kierownictwem prof. Adama Wierzbickiego w Polsko-Japońskiej Akademii Technik Komputerowych, skłaniają mnie do przyjęcia, że ważnym komponentem wiarygodności są nie tylko właściwości językowe wypowiedzi, ale też ich zgodność z pewnym kanonem wiedzy. Przy takim rozumieniu tego pojęcia próba mówienia o ocenie podobieństwa wiarygodności dwóch wypowiedzi staje się o wiele trudniejsza. Jednakże, nawet jeżeli taką funkcję podobieństwa (zbudowaną na bazie modeli wektorowych) zaproponować i wdrożyć, to wciąż może ona być podatna na ograniczenia wynikające z właściwości użytych narzędzi.

Otóż, jak już wspomniałem, modele wektorowe wcześniej uczą się znaczenia słów w oparciu o to, w jakim kontekście się pojawiają. Oznacza to, że będą raczej reprezentować sposób użycia danych słów i fraz w kontekstach reprezentatywnych dla tre-

ści na których się uczyły. Badania wskazywały, że może to prowadzić do reprodukcji uprzedzeń (np. seksistowskich) już istniejących w danych społecznościach⁹. Być może zatem taki wspomagany *fakt-checking* nie jest wystarczającym narzędziem do eliminacji problemu fejków. Być może powinniśmy w pełni zautomatyzować ten proces i wykorzystać zdobyte uczenia maszynowego?

AUTOMATYCZNA KLASYFIKACJA

Uczenie maszynowe zakłada, że program w oparciu o przygotowany uprzednio algorytm uczenia maszynowego oraz dostępne dane treningowe będzie w stanie sam nauczyć się, jak wykonywać zadanie, które przed nim stawiamy. Uczenie maszynowe, w odróżnieniu od sztywnego pisania instrukcji dla programu, ma ogromną przewagę w sytuacjach kiedy zadanie jest zmienne, i w zależności od kontekstu może wymagać innych działań.

Obecnie dominującą formą uczenia maszynowego są sztuczne sieci neuronowe – algorytmy, które w swojej strukturze wzorowane są na tkankach nerwowych obecnych w przyrodzie. Od czasu ich pierwszego wprowadzenia rozwinęło się wiele typów takich sieci dostosowanych do różnych typów zadań. Nic zatem nie stoi na przeszkodzie, żeby jedną z nich wykorzystać, aby nauczyła się rozpoznawać fejki. Takie rozwiązania już się faktycznie testuje, głównie z wykorzystaniem tzw. rekurencyjnych sieci neuronowych albo sieci LSTM¹⁰.

Takie sieci prawidłowo zastosowane powinny dawać wymierne efekty i na ogół w laboratoriach takie wyniki dają.

Jednakże kiedy te algorytmy mają już być zastosowane w kontekście mediów społecznościowych, ich skuteczność spada. Sieci LSTM są bardzo elastyczne i potrafią dopasować się do nowych danych, tym niemniej głównym problemem

Obecnie dominującą formą uczenia maszynowego są sztuczne sieci neuronowe – algorytmy, które w swojej strukturze wzorowane są na tkankach nerwowych obecnych w przyrodzie.

Nic zatem nie stoi na przeszkodzie, żeby jedną z nich wykorzystać, aby nauczyła się rozpoznawać fejki. Takie rozwiązania już się faktycznie testuje, głównie z wykorzystaniem tzw. rekurencyjnych sieci neuronowych albo sieci LSTM.

nie jest sama struktura sieci, co zadanie jakie mają wykonać. O części ograniczeń wspomniałem już w poprzednich akapitach (zależność od zbioru treningowego, istotność tego, kto zbiór treningowy przygotowuje – tym samym decydując, co jest wiarygodne, a co nie), jednakże od strony ludzkiej, głównym problemem na który należy zwrócić uwagę przy użyciu uczenia maszynowego jest sytuacja kiedy model daje dobre wyniki i w laboratorium, i w praktyce. Jeżeli do czegoś takiego dochodzi, administrator modelu powinien dbać o to, by monitorować to, jak model się myli oraz czy z czasem nie pojawiają się systematyczne błędy.

PODSUMOWANIE

W tym krótkim eseju próbowałem przedstawić główne strategie walki ze zjawiskiem *fake news* wykorzystywane przez badaczy z dziedziny informatyki. Większość czytelników pewnie dosyć szybko zauważy, że byłem dosyć krytyczny wobec każdej z nich, tym samym dając podstawę do stwierdzenia, że nie uważam, by takie narzędzia mogłyby kiedykolwiek być zastosowane w praktyce.

16 Nie jest to w pełni prawdą. Moje krytyczne podejście do tych technologii wynika nie z faktu, że są zbyt proste albo przeciwnie – zbyt złożone obliczeniowo. Starałem się tylko wskazać ograniczenia, w jakich te technologie obecnie funkcjonują oraz przybliżyć czytelnikom jakie systemowe rozwiązania są proponowane przez specjalistów. Jednakże te ograniczenia są tylko kwestiami technologicznymi. O wiele poważniejszym problemem w kontekście *fake news* nie jest to, jak wykrywamy fejki, ale to, kto decyduje o tym, co fejkem jest.

Każde społeczeństwo, nawet to najbardziej racjonalne i modernistyczne ma swoje mity. I nie ma w tym nic złego, bo jak twierdził Durkheim bądź Eliade – w mitach i dogmatach społeczeństwo odtwarza wyobrażenie o sobie samym, idealnym. Jednym z takich mitów naszego świata jest wiara w maszynę. Nie jest to oczywiście przeświadczenie wyrażane świadomie, raczej obecne jako założenie, że narzędzie zautomatyzowane jest lepsze, szybsze, tańsze i, co najważniejsze, wolne od ludzkich błędów i wad. Idealny system do wykrywania fejków wpisuje się w taki mit – niezależna od naszych upodobań i sympatii maszyna, która wskaże nam, co jest wiarygodne, a co nie. Fakt, że to, co wiarygodne, będzie zgodne z tym, co już za prawdę uznaliśmy, będzie tylko statystyczną anomalią.

Wskazanie ograniczeń oraz roli czynnika ludzkiego w technologiach limitujących propagację fejków było głównym celem tego krótkiego eseju. To prawda: *fake newsy* są szkodliwe, rozbijają spójność społeczną, są narzędziem do najbardziej prymitywnej gry politycznej w rękach cyników, którym nie przeszkodzi krew na rękach, bo przecież oni to „ci dobrzy” (albo po prostu chcą władzy). W końcu też *fake newsy* niszczą naszą zdolność odbioru rzeczywistości, gdyż jeszcze bardziej zamykają nas w bańce tego, co chcemy by prawdą było.

Jednakże nie oznacza to, że każde działanie w nie wymierzone będzie automatycznie korzystne bądź służące prawdzie. Platformy *fact-checkingowe* mogą równie dobrze zaznaczać jako fejki treści, które szkalują wybrany autorytet, algorytmy uczenia maszynowego wytrenują się rozpoznawać fałszywki na temat tylko wybranej mniejszości (gdyż tylko ta jedna mniejszość jest ofiarą QED); w końcu możemy połączyć wszystkie te narzędzia, by nikt już nigdy nie podważał naszej przewodniej roli w państwie...

Koniec końców, najważniejsza w zwalczaniu fejków jest nasza własna postawa jako odbiorców mediów. Wspomniałem na samym początku, kiedy powoływałem się na podobieństwa między teoriami spiskowymi a internetowymi fałszywkami, że jednym ze źródeł popularności takiej formy narracji oraz organizacji wiedzy jest fakt, że są one „fajne”. Wtedy nie rozwinąłem, co za tym pojęciem miałyby się kryć.

Poprzez użycie słowa „fajne” miałem na myśli fakt, że tego typu treści wpisują się w pewną wizję świata, którą już mamy. Nie musimy w pełni się z nią zgadzać, nie musimy akceptować jej w sposób świadomy, ale odpowiada ona temu, czego oczekujemy. Oczywiście pierwszym skojarzeniem będą fałszywki, które potwierdzają nasze poglądy – tutaj mechanizm jest zrozumiały jako element efektu potwierdzenia (*confirmatio bias*). Jednakże nie wyczerpuje to możliwości psychologicznego wpływu *fake news* na odbiorców. Mniej oczywistym i przez to bardziej niebezpiecznym ich przykładem będą fałszywki, które mają wywołać w odbiorcy strach, wzmóc poczucie zagrożenia. Oznacza to, że łańcuszek o zbrodni dokonanej przez migrantów na niewinnej kobiecie przejdzie bez problemu przez nasz filtr sceptycyzmu, gdyż narracja „kobiet-ofiar i obcych-oprawców” jest w nas zakodowana jako „prawdopodobne” i wpisuje się w realne zagrożenie. Kiedy odczytamy nagłówek o petycji karzącej więzieniem za obchodzenie święta Halloween, nie sięgamy głębiej by sprawdzić prawdziwość takiej informacji, wszak to „prawdopodobne”, że taka petycja się pojawi.



Dopóki będziemy czytać wyłącznie nagłówki, czytać tylko to, z czym już się zgadzamy, uważać, że posiadamy monopol na prawdę i dawać naszym emocjom pierwszeństwo w reakcji, *fake news* będą miały po swojej stronie najsilniejszy z argumentów – opłaczalność wynikającą z kolejnych kliknięć, wyświetleń i reakcji.

Dopóki będziemy czytać wyłącznie nagłówki, czytać tylko to, z czym już się zgadzamy, uważać, że posiadamy monopol na prawdę (w odróżnieniu od tych oczadziawych, zmanipulowanych „innych”) i dawać naszym emocjom pierwszeństwo w reakcji, *fake news* będą miały po swojej stronie najsilniejszy z argumentów – opłaczalność wynikającą z kolejnych kliknięć, wyświetleń i reakcji.

BIBLIOGRAFIA

- [1] Pennycook G., Bear A., Collins E. T., Rand D. G. (2020). *The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings*. Management Science.
- [2] Zou J. & Schiebinger L. (2018). *AI can be sexist and racist—it's time to make it fair*.
- [3] Bolukbasi T., Chang K. W., Zou J. Y., Saligrama V., Kalai A. T. (2016). *Man is to computer programmer as woman is to homemaker? Debiasing word embeddings*. In *Advances in neural information processing systems* (pp. 4349-4357).
- [4] Shu K., Sliva A., Wang S., Tang J., Liu H. (2017). *Fake news detection on social media: A data mining perspective*. ACM SIGKDD explorations newsletter” 19(1), 22-36.
- [5] Ruchansky N., Seo S., Liu Y. (2017, November). *Csi: A hybrid deep model for fake news detection*. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).

PRZYPISY KOŃCOWE

- [1] W ZSRR funkcjonowała gra słowna nawiązująca do największych sowieckich gazet: „Prawdy” i „Izwestii” (ros. „Wiadomości”) - W „Pravdie” niet izvietii a w „Izvestiach” niet pravdy.
- [2] Czasami bot może być nieodróżnialny od sztampono piszącego stażysty.
- [3] Tandoc Jr, Edson C., Zheng Wei Lim, and Richard Ling. *Defining “fake news” A typology of scholarly definitions*, “Digital journalism” 6.2 (2018): 137-153.
- [4] Uscinski J.E., Klofstad C., Atkinson M.D. (2016) *What drives conspiratorial beliefs? The role of informational cues and predispositions*. Polit Res Q 69: 57–71.
- [5] <https://demagog.org.pl/>
- [6] <https://www.politifact.com/>
- [7] Teoria starożytnych kosmitów jest alternatywną teorią dziejów powstałą w latach 70. i bazującą przede wszystkim na książkach Ericha von Danikena oraz Zacharia Stichina. Teoria starożytnych kosmitów postuluje, że w czasach starożytnych rasa (bądź rasy) inteligentnych istot pozaziemskich wylądowały na Ziemi i przyczyniły się do narodzin licznych starożytnych cywilizacji takich jak Sumer lub Egipt. Obecnie głównymi źródłami propagacji tej teorii jest Internet oraz emitowany na History Channel program „Ancient Aliens”.
- [8] <https://www.cs.york.ac.uk/semEval-2012/index.html>
- [9] Bolukbasi T., Chang K. W., Zou J. Y., Saligrama V., Kalai A. T. (2016). *Man is to computer programmer as woman is to homemaker? Debiasing word embeddings*. In *Advances in neural information processing systems* (pp. 4349-4357).
- [10] Ruchansky N., Seo S., Liu Y. (2017, November). *Csi: A hybrid deep model for fake news detection*. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).

BEST TOOLS FOR DETECTING FAKE NEWS (AND WHY THEY DO NOT WORK)

ENGLISH SUMMARY

The article provides a review of various technologies used in detecting fake news online. It also provides discussion of sociological and psychological factors which influence the popularity of fake news. Finally a discussion of political consequences of the use of said technologies is provided.

Keywords: fake news, computational social science, natural language processing, machine learning

Bartłomiej Balcerzak

Polsko-Japońska Akademia Technik Komputerowych



nr ORCID **0000-0003-0881-5362**

Absolwent socjologii na IS UW oraz Informatyki Społecznej w Polsko-Japońskiej Akademii Technik Komputerowych, gdzie również prowadzi zajęcia dydaktyczne z zastosowania analiz statystycznych, uczenia maszynowego oraz Big Data. Współpracuje również z SWPS. Członek zespołu badawczego skupionego na analizie wiarygodności treści medycznych w Internecie, oraz kolektywu naukowego BiTech.