

POLITYKA

Jacek Żakowski, 17 stycznia 2017

Jak politycy wykorzystują algorytmy, by zdobyć nasze głosy

Internetowy Frankenstein

Prof. Michał Kosiński o tym, jak nowoczesne metody marketingowe pozwalają wpływać na polityczne decyzje każdego człowieka.

Artykuł dostępny tylko dla abonentów Polityki Cyfrowej

Jacek Żakowski: – **To pan rozwalil świat!**

Michał Kosiński: – W jakim sensie?

Przez pana Trump będzie prezydentem.

Nigdy nie pracowałem dla Trumpa ani dla żadnego innego polityka.

Wystarczy, że pana metoda pracowała. Podobno za sukcesem Brexitu i Donalda Trumpa stoi firma Cambridge Analytica, która zaskoczyła konkurencję, stosując innowacyjny psychograficzny mikrotargetingowy marketing oparty na pańskich badaniach.

Jeżeli to poprawi humor liberałom na całym Zachodzie, mogę wziąć tę odpowiedzialność na siebie. Chociaż od dawna w kółko piszę i mówię, że również ta wiedza może być wspaniale użyteczna, ale może też być groźna, jeśli się jej użyje w złym celu.

Jak oni to zrobili?

Najpierw trzeba powiedzieć, jak się to robiło dotychczas. Bo tradycyjny marketing, także polityczny, oparty był na danych demograficznych. Można było kierować komunikat do kobiet, do młodych, do zamożnych, do wykształconych, do ludzi z dużych miast albo ze wsi. To są wielkie i zróżnicowane grupy.

Albo do młodych zamożnych kobiet z wielkich miast.

Nawet jeśli się te cechy krzyżowało, były to wielkie i wciąż mocno zróżnicowane grupy. To przypominało strzelanie do stada wróbli ze śrutu. Któryś pocisk któregoś ptaka trafił. A większość odlatywała. Bo młode, wykształcone kobiety z wielkiego miasta statystycznie mają pewne wspólne ważne marketingowo cechy, ale jedne z nich wołają, kiedy się do nich mówi szybko i wesoło, inne kiedy mówi się wolno i poważnie. Jedne lubią muzykę soulową, a inne hard rock. Jedne wołają kolor zielony, a drugie brązowy. Jedne lubią mężczyzn dużych i umięśnionych, a inne wrażliwych i delikatnych.

I jedne głosują na PO, a drugie na PiS.

No właśnie. Grupy demograficzne są wewnątrz zróżnicowane pod wieloma względami. Jeżeli się do takiej grupy kieruje ten sam komunikat, na przykład reklamując coś przy pomocy wesołego umięśnionego mężczyzny ubranego w brązy i szybko mówiącego na tle romantycznej muzyki, to dla części będzie to bardzo przekonujące, a innych rozdrażni. Mikrotargetowanie to wysyłanie zróżnicowanych komunikatów precyzyjnie dostosowanych do konkretnej osoby.

Jeśli lubisz umięśnionego i romantyczną muzykę, dostajesz taką reklamę proszku, samochodu albo polityka, a jak lubisz umięśnionego i heavy metal, to będzie cię przekonywał mięśniak z ciężkim brzmieniem w tle.

Dokładnie. To działa nieporównanie skuteczniej niż na przykład jednakowe billboardy przy drodze albo internetowe reklamy trafiające do wszystkich młodych czy wszystkich, którzy interesowali się jakimś tematem.

Jak bardzo skuteczniej?

Według moich badań skuteczność psychograficznego mikrotargetowania może być nawet kilkunastokrotnie większa od skuteczności reklam wysyłanych według kryteriów demograficznych. Nie badałem różnicy w przypadku przekazu politycznego, ale efekty są prawdopodobnie podobne.

Czyli jeżeli wesoly umięśniony mężczyzna ubrany w brązy i szybko mówiący na tle romantycznej muzyki będzie przekonywał kobietę, która to wszystko lubi, do Trumpa albo do Brexitu, będzie kilka razy bardziej skuteczniejszy niż aktor z tradycyjnej reklamy?

Zwłaszcza jeśli będzie używał argumentów, na które ona jest wrażliwa. Bo do tego samego można przekonywać przy pomocy bardzo różnych argumentów. Osoby wrażliwe moralnie można do Trumpa przekonywać, mówiąc, że establishment jest przegniły i zdemoralizowany; drobny biznes, który źle sobie radzi, można pozyskać, obiecując twardą walkę z Wall Street; a ludzi, których dzieci mają problem z narkotykami, obiecując programy odwykowe.

Tak Cambridge Analytica miało działać w kampanii dotyczącej Brexitu i w amerykańskich wyborach. Podobno wolontariusze Trumpa chodzący po domach mieli w smartfonach program, który po wklepaniu adresu podawał im dokładne profile każdego domownika. „Mąż jest nasz, a żona głosuje na Clinton. Na nich szkoda czasu. Ale dorosła córka, która po rozwodzie wróciła do rodziców, waha się, na kogo głosować, więc warto z nią porozmawiać. Jest wrażliwa na argumenty moralne, martwi się o to i o to, lubi to i to”. Prawdopodobne?

To staje się nowym standardem w marketingu, biorąc pod uwagę, jak wiele śladów zostawiamy po sobie w internecie i jak dokładne są już modele komputerowe. Moje badania w Instytucie Psychometrii Uniwersytetu w Cambridge – z którym Cambridge Analytica nie ma nic wspólnego – wykazały, że w internecie ludzie zachowują się spójnie z tym, jacy są w realu. Obserwując zachowania ludzi w sieci, można dokładnie przewidzieć ich opinie, postawy czy cechy intymne. Zamiast korzystać z ankiet, gdzie przecież odpowiedzi nie muszą być szczere, można więc badać cyfrowe ślady, takie jak maile, historie wyszukiwania, odwiedzone strony itd. Z moich badań wynika, że na podstawie 10 facebookowych lajków sztuczna inteligencja może ocenić osobowość człowieka trafniej niż kolega z pracy; na podstawie 70 lajków – oceni lepiej niż przyjaciel; na podstawie 150 – lepiej niż rodzice; a na podstawie 250 – lepiej niż partner życiowy. Może być tak, że algorytm zrozumie pana lepiej niż pan sam. Teraz na podstawie 100 do 200 lajków jesteśmy już w stanie określić poglądy polityczne dowolnej osoby z paroprocentowym marginesem błędu. I to po odrzuceniu wszystkich lajków dotyczących polityki.

Nie używam Facebooka, więc jestem bezpieczny.

Proszę się nie ludzić! Może pan nie mieć Facebooka, ale zostawia pan w sieci masę śladów. Dane Twittera są publicznie dostępne.

Twittera też nie mam.

Nie szkodzi. Ma pan konto w banku. I kartę kredytową albo lojalnościową pan ma. Ma pan też smartfon. On śledzi lokalizację. Pana operator zapewne nie zapisuje treści pana rozmów, ale wie, jak długo i z kim rozmawia pan przez telefon. Algorytm może też czytać pana esemesy. Takie dane można też łatwo kupić. Bazy zawierające szczegółowe dane milionów ludzi są na sprzedaż na internetowych rynkach baz danych. Zawierają listy odwiedzanych stron, profile społecznościowe, wiadomości z Twittera, listy pana znajomych, informacje o pana kredytach, o książkach, jakie pan czyta, miejscach, gdzie pan spędza wakacje i które wybiera pan w samolocie. Dane o zakupach zbierane za pomocą kart lojalnościowych i kredytowych także są na sprzedaż. Facebook, na przykład, kupuje takie dane, aby optymalizować reklamy, które pokazuje swoim użytkownikom.

Ale te dane nie zawierają informacji o tym, czego się boję, jakie mam marzenia, poglądy, osobowość albo czy waham się w sprawie Brexitu.

Bardzo się pan myli. Algorytmy mogą przetworzyć listy zakupionych przez pana towarów czy wysłuchanych piosenek w bardzo dokładny profil psychologiczny. Ludzkie zachowania i preferencje nie są przypadkowe – to, jakie kupuje pan buty, jest powiązane z tym, jaką pan ma osobowość czy poglądy polityczne. Taka zależność może być bardzo słaba – dla człowieka niezauważalna, ale algorytm umie ją wychwycić. Być może konserwatywni ludzie kupują adidasów o setną część procentu częściej niż liberałowie. Sama w sobie taka informacja nic nie znaczy. Algorytm może jednak połączyć setki czy tysiące takich małych zależności w bardzo dokładną diagnozę nie tylko poglądów politycznych konkretnej osoby, ale też bardzo intymnych cech, przekonań czy postaw.

Jak pan na to wpadł?

Zaczęło się od kwestionariusza mierzącego tak zwaną wielką piątkę cech osobowości, który kilka lat temu umieściliśmy w sieci. Każdy mógł go sobie bezpłatnie i anonimowo wypełnić, żeby dowiedzieć się, jakiego typu ma osobowość. Kilka milionów osób z tego skorzystało. Dzięki temu zebraliśmy być może największą w historii bazę danych łączącą osobowości ludzi i ich cyfrowe ślady. Szybko zauważyliśmy, iż zamiast pytać ludzi o to, jak często czytają książki lub czy są szczęśliwi, możemy policzyć, jak często odwiedzają strony dotyczące książek albo jak często używają pozytywnych słów w swoich wiadomościach. Okazało się, iż wyniki badania cyfrowych śladów są często bardziej wiarygodne niż oparte na kwestionariuszach. Możemy w ten sposób poznać osobowość, inteligencję, zainteresowania.

I także jak ktoś głosuje w wyborach czy referendum albo dlaczego się waha i co go przekona?

Dokładnie tak. Dzięki Big Data, czyli wciąż mnożącym się gigantycznym bazom obejmującym informacje o milionach ludzi, zdarzeń i procesów, z korelacji danych pozornie zupełnie obojętnych – na przykład tembru głosu, jakim pan rozmawia przez telefon, liczby kart kredytowych, podróży, liczby otrzymanych i pisanych maili, sklepów, które pan odwiedza itd. – można wnioskować z bardzo dużą pewnością. Nie tylko, jakie pan ma poglądy, ale też czy pańscy rodzice się rozwiedli, w jakim stanie jest pański obecny związek, jakie ma pan zwyczaje i marzenia seksualne, jaka jest pańska postawa religijna, kiedy podejmie pan decyzję wyborczą albo zakupową, czego się pan boi, na co pan liczy, czy używa pan narkotyków i jakich, czy jest pan szczęśliwy i inteligentny, jakie kolory, muzykę pan lubi, jakie argumenty do pana trafiają i jakie wartości są dla pana najważniejsze. Wszystkie te cechy mogą być ustalone na podstawie informacji pozornie obojętnych.

Nie mogę się przed tym ukryć?

Lepiej na to nie liczyć. Musi pan przyjąć do wiadomości, że jeśli ktoś chce się czegoś o panu dowiedzieć, to się dowie. Jeżeli nie wprost, to analizując różne korelacje.

A jak będę zostawiał fałszywe tropy?

Algorytm bez trudu to wykryje. I szybko ustali, jakie cechy mają osoby kombinujące podobnie jak pan. Przy dziesiątkach czy setkach milionów analizowanych osób trudno być oryginalnym. W tej skali nawet najbardziej ekstrawagancka postawa ma zakres w tradycyjnych kategoriach masowy, więc umożliwi algorytmowi dokładne rozszyfrowanie każdego z nas dzięki korelacjom danych

dotyczących milionów innych osób. Tu powstaje moralne ryzyko, przed którym wciąż ostrzegam. Bo kiedy dawniej przychodził do pana ankieter, mógł go pan przepędzić albo oszukać i w ten sposób ukryć, co pan myśli. Metody, którymi się od kilku lat zajmuję, sprawiają, że dziś już pan takiej szansy nie ma.

Skąd ten algorytm wie, że liberalowie częściej chodzą w jakichś butach albo jeżdżą gdzieś na wakacje?

Nasz algorytm sam się tego uczy, obserwując świat. To jest bardzo proste. Śledzi miliony ludzi. Jakąś część z nich udaje mu się zidentyfikować wprost – przez informacje o przynależności partyjnej, wejściach na strony sklepów albo przez wpisy czy lajki na Facebooku, Twitterze lub forach. Sprawdza, jakie inne zachowania łączą te osoby, a potem znajduje osoby, które się zachowują podobnie.

Jak to się stało, że Cambridge Analytica użyło tego modelu, żeby wesprzeć Brexit i Trumpa?

Ja się o tym dowiedziałem z prasy. Ich firma matka, czyli SCL, proponowała mi wcześniej współpracę. Odmówiłem, kiedy się zorientowałem, że chodzi o uzyskanie wpływu na wyniki wyborów. Ale wyniki naszych badań są powszechnie dostępne, więc mogli na ich podstawie zbudować własne modele. Wiele firm, a nawet pojedynczych osób może te metody skutecznie stosować.

Są takie komputery, które potrafią w ten sposób roznegliżować na przykład każdego Polaka z imienia i nazwiska?

Ja pracuję na bazach danych obejmujących dziesiątki milionów ludzi, a większość badań mogę prowadzić na zwykłym laptopie. Technicznie to nie jest żaden problem. Kluczowe są bazy danych i materiał pozwalający na standaryzację w odniesieniu do jakichś szczegółowych problemów. Cambridge Analytica twierdzi, że uzyskali je, umieszczając dobrowolnie wypełniane ankiety w internecie. To jest bardzo prawdopodobne. Jako podstawy do profilowania można też użyć sondaży internetowych, których są tysiące na wielu portalach. Jeżeli weźmie w nich udział wystarczająca grupa, badając prawidłowości, można bardzo precyzyjnie psychograficznie profilować każdego.

Tych, którzy nie wzięli udziału, też?

Na tym ta metoda polega.

Mam nadzieję, że w Polsce to jeszcze tak nie działa.

Działa tak samo jak wszędzie. W dużym stopniu właśnie tak na całym świecie funkcjonują cyfrowe platformy marketingowe. Facebook czy Google działają bardzo podobnie niezależnie od kraju. Nawet w przypadku cyfrowej telewizji reklamy dobierane są indywidualnie dla każdego widza – zapewne ogląda pan inne reklamy niż pana sąsiad.

Nie wiem. Ale trochę mnie to wszystko przeraża.

To może przerażać. Ale też fascynuje. Bo to jest też wspaniałe. Niech pan pomyśli, ile osób było dotąd faktycznie wykluczonych z demokratycznej polityki, bo nic z tych rzeczy, o których politycy mówili, ich nie interesowało. Tacy ludzie byli wyłączeni z procesu politycznego. Nie czytali gazet, nie chodzili na spotkania z politykami, nie głosowali. Teraz do każdego z nich polityk może dotrzeć z przesłaniem, które ich zainteresuje. Co każdego takiego człowieka poruszy, można się dowiedzieć dzięki algorytmom, a dzięki mikrotargetingowi polityk może dotrzeć do niego z precyzyjnie dobranym przekazem.

Jakoś się tym nie umiem zachwyścić.

A tym, że w taki sam sposób wydawca może do pana dotrzeć z informacją o książkach, które pana zainteresują, a nie słyszał pan o ich istnieniu?

Wydaje mi się bardziej prawdopodobne, że taki algorytm, który zna mnie na wylot, skutecznie namówi mnie do kupienia rzeczy, których mi nie potrzeba, albo złapie mnie na haczyk politycznej obietnicy, której nikt nie dotrzyma.

Dlatego wciąż przestrzegam, że istnieje ryzyko nadużycia. I bez tego polityka jest przecież pełna niespełnionych obietnic, a reklama uwodzi nas i skłania do absurdalnych zakupów. Sieć ma jednak zasadniczą przewagę nad realem – jest nią sprzężenie zwrotne. Politycy czy producenci znają nas lepiej niż kiedykolwiek wcześniej, ale my także znamy ich lepiej niż kiedykolwiek. W przeszłości mógł pan sprzedawać ludziom buble lub obiecywać gruszki na wierzbie i dużo czasu potrzeba było, by stało się to powszechnie znanym faktem. Teraz kilka negatywnych komentarzy na Allegro może skończyć karierę największego sprzedawcy. A tweet nastolatka źle potraktowanego przez dyrektora szkoły może skończyć dyrektorską karierę.

OK, ale tymczasem, między innymi dzięki metodzie zastosowanej przez Cambridge Analytica, Anglia wychodzi z Unii, a Trump właśnie wprowadza się do Białego Domu. O kłamstwach i nadużyciach, które stały za tymi kampaniami, dowiadujemy się dopiero, kiedy kłamki zapadły.

Nie wiemy, jaka dokładnie była skuteczność Cambridge Analytica, bo oni żadnych danych nie ujawnili. Ale mogła być spora.

Wygląda na to, że niewielka przewaga, jaką uzyskali zwolennicy Brexitu i Trumpa, wynikała z zastosowania mikrotargetingu, bo druga strona spóźniła się na tę rewolucję.

Nie wiemy, czy przeciwnicy Brexitu i Trumpa nie stosowali psychograficznego mikrotargetowania. Nikt się do tego nie przyznał, ale to o niczym nie świadczy. Ze sztabem Clinton współpracował przecież m.in. słynny Eric Schmidt z Google. Mnie niepokoi coś zupełnie innego. To, że profile naszych intymnych cech robione są za naszymi plecami. Nikt nas o tym nie informuje i nikt nie pyta, czy może w kontakcie z nami używać informacji pochodzących z psychograficznych profili internetowych.

Bo gdyby ludzie zdawali sobie sprawę, że byli profilowani, staliby się bardziej czujni, kiedy ktoś próbuje ich w oparciu o profil do czegoś przekonać.

Oczywiście. Chociaż wielu zgodziłoby się na to. Ja, na przykład, bardzo się cieszę, że algorytm Amazona zna mnie już na wylot, bo dzięki temu polecił mi wiele wspaniałych książek, których sam bym nigdy nie znalazł. Ale więcej przejrzystości i szacunku dla prywatności użytkownika byłoby nie tylko fair, lecz budowałoby też zaufanie. Firmy powinny się tym bardziej przejmować, bo brak zaufania ze strony użytkowników w dłuższej perspektywie źle biznesowi wróży.

Można coś zrobić, żebyśmy jako jednostki i jako społeczeństwa nie stawali się bezbronnymi ofiarami takich manipulacji?

Jedyną radą jest chyba edukacja. Ludzie wykształceni, odcytani, poinformowani stają się bardziej krytyczni, więc trudniej jest nimi manipulować. Trzeba ludzi ostrzegać. Trzeba dyskutować o takich zdarzeniach, jak wykorzystanie przez Trumpa psychologicznych profili uzyskanych bez wiedzy i zgody ludzi, którzy byli targetowani przez jego marketing. Ale dużo większym problemem jest to, że taka technologia może być używana przez rządy i instytucje państwowe. Rządy państw niedemokratycznych także mogą korzystać z takich technologii, by poznawać poglądy albo orientację seksualną swoich obywateli. A nasz świat nie zawsze musi być taki wolny jak dziś.

Więc może trzeba pomyśleć o regulacjach?

Można sobie wyobrazić prawo bardziej ograniczające to, co firmy mogą z naszymi danymi robić. Trzeba takie prawa przemyśleć i wdrożyć. Ale pokusa, by takie prawo naruszać, będzie ogromna. Pewnie większa niż w przypadku handlu narkotykami. Powstanie wielka szara strefa i potężne podziemie.

I będziemy mieli kolejną „wojnę z narkotykami”, która pewnie będzie miała bardzo podobne skutki.

Albo gorsze. To jest duże ryzyko, bo algorytmy się uczą i są z czasem coraz bardziej dokładne, a nasze życie coraz bardziej przenosi się do świata cyfrowego. Dlatego nawet jeżeli uchwalimy najdoskonalsze prawo, prywatności będzie coraz mniej. Najlepszym sposobem obrony jest budowanie społeczeństw bardziej wykształconych, otwartych, tolerancyjnych i demokratycznych. Żeby zanik prywatności stawał się mniej groźny i żeby ludzie lepiej umieli się bronić przed manipulacjami, których narzędzia będą coraz doskonalsze. Ale jeżeli my będziemy mądrzejsi, te doskonalsze narzędzia nie muszą być groźniejsze – bez względu na to, kto będzie się nimi posługiwał – Trump, wielka korporacja czy jakiś satrapa.

rozmawiał Jacek Żakowski

Michał Kosiński ukończył psychologię społeczną na SWPS. Pracował i doktoryzował się na Uniwersytecie Cambridge, gdzie był też wicedyrektorem Centrum Psychometrycznego. Pracuje na Uniwersytecie Stanforda oraz jako profesor w Szkole Biznesu. W 2013 r. IBM i DataIQ zaliczyły go do 50 najbardziej wpływowych osób w świecie Big Data. W 2016 r. został uznany za Wschodzącą Gwiazdę przez Association for Psychological Science.

Serwis redaguje zespół POLITYKA.PL

Wykonanie Javatech | Prawa autorskie © POLITYKA Sp. z o.o. S.K.A.