FUNDACJA PANOPTYKON

Śledzenie i profilowanie w sieci

Jak z klienta stajesz się towarem

Katarzyna Szymielewicz, Karolina Iwańska

Styczeń 2019

Wprowadzenie

Spór o prywatność w sieci rozbija się o to, jaki model finansowania treści dostępnych w sieci jesteśmy gotowi wspierać i czy potrafimy go sobie wyobrazić bez targetowanej reklamy. Większość wydawców mediów internetowych i reklamodawców tego nie potrafi, więc stawia użytkowników przed nieuczciwym wyborem: albo pozwolicie się śledzić i profilować, albo odetniemy was od naszych usług.

Nie negujemy tego, że media mają prawo i potrzebę zarabiać na serwowanych treściach. Stawiamy jednak pytanie o to, czy rynek reklamowy musi się rozwijać w oparciu o nieprzejrzyste i nieetyczne praktyki, takie jak śledzenie ludzi bez ich wiedzy i zgody.

> **Głównym celem wydawców i reklamodawców jest wysoka klikalność banerów reklamowych, a to wymaga śledzenia i profilowania. Dla użytkowników ten model komercjalizacji danych oznacza realne ryzyka. Nie chodzi tylko o to, czy kupią kolejną parę butów albo (jeszcze) nowszy telefon. Stawką w tej grze jest kontrola nad ich cyfrowym „ja".**

Techniki śledzenia i profilowania stają się coraz bardziej inwazyjne, nie oszczędzając żadnej sfery życia prywatnego. Dane dotyczące zdrowia (także intymnego czy psychicznego), sytuacji finansowej, pochodzenia etnicznego, relacji osobistych, nałogów, słabości, marzeń i aspiracji miliardów ludzi są zbierane lub generowane (na zasadzie predykcji) oraz integrowane w sposób, który nie uwzględnia możliwych ryzyk. Kojarzeniem tego typu danych poza kontrolą ludzi, których one dotyczą, są zainteresowani nie tylko ubezpieczyciele czy potencjalni pracodawcy, ale coraz częściej także partie konkurujące w wyścigach wyborczych.

—

Żeby pokazać skalę, metody i techniki śledzenia wykorzystywane w usługach internetowych przekopaliśmy się przez badania, raporty organizacji społecznych i branżowych. Odbyliśmy też długie rozmowy z przedstawicielami firm pełniących różne funkcje w ramach reklamowego ekosystemu. Bez ich głosu ta historia byłaby niepełna. Nadal mamy świadomość, że pokazujemy tylko wierzchołek góry lodowej. Większość procesów, o których piszemy, odbywa się bowiem w cieniu, a odtworzenie ich przebiegu i oddziaływania na użytkowników sprawia trudność nawet technicznym ekspertom.

Traktujemy tę publikację jako kolejny – lecz z pewnością nie ostatni – krok do ich zmapowania i opisania[1].

Za wsparcie merytoryczne w pracach nad raportem dziękujemy Maciejowi Zawadzińskiemu – współzałożycielowi i Prezesowi firmy Piwik PRO oraz Michałowi Czaprackiemu i Karolinie Gębce z firmy Signiroad.

Gorące podziękowania za opracowanie danych do infografiki o cyfrowym profilu oraz wykresów obrazujących dane otrzymane od jednego z portali składamy naszemu wolontariuszowi Marcinowi Antasowi (jego portfolio publiczne).

Kamilowi Śliwowskiemu (Otwarte Zasoby) dziękujemy za opracowanie grafiki o cyfrowym profilu i nieocenione sugestie, które pomogły nam w pisaniu o śledzących aplikacjach i smart urządzeniach.

Autorem infografik o przebiegu aukcji na giełdzie reklam jest Jacek Rakiej (jego portfolio).

[1] Ta publikacja stanowi zaktualizowaną i uzupełnioną wersją raportu na temat śledzenia i profilowania w sieci, opublikowanego we wrześniu 2017 r., https://panoptykon.org/sites/default/files/publikacje/sledzenie_i_profilowanie_w_sieci_scenariusze_po_reformie_ue_wrzesien_2017.pdf.

Komercyjny Internet

Rzut oka z perspektywy 2019 r.

W 2019 r. obchodzimy 25-lecie pierwszego zamieszczonego w Internecie baneru reklamowego. Od tego czasu dane osobowe pod różnymi postaciami stały się nową walutą. Szacuje się, że wartość rynkowa ogółu danych przetwarzanych online na terenie Unii Europejskiej osiągnie do 2020 r. wysokość 739 miliardów euro[2]. Od kilku dekad różne branże (handlu detalicznego, turystyczna, dóbr konsumpcyjnych, mediów, telekomunikacyjna, bankowa, ubezpieczeniowa i inne) gromadzą i wykorzystują informacje o swoich klientach, a częściowo także wymieniają się nimi. Firmy nauczyły się skutecznie wykorzystywać techniki śledzenia w sieci do identyfikowania, zdobywania i przywiązywania do siebie atrakcyjnych klientów, wyliczania ich wartości oraz do „efektywnego inwestowania środków" w „najbardziej dochodowe grupy klientów".

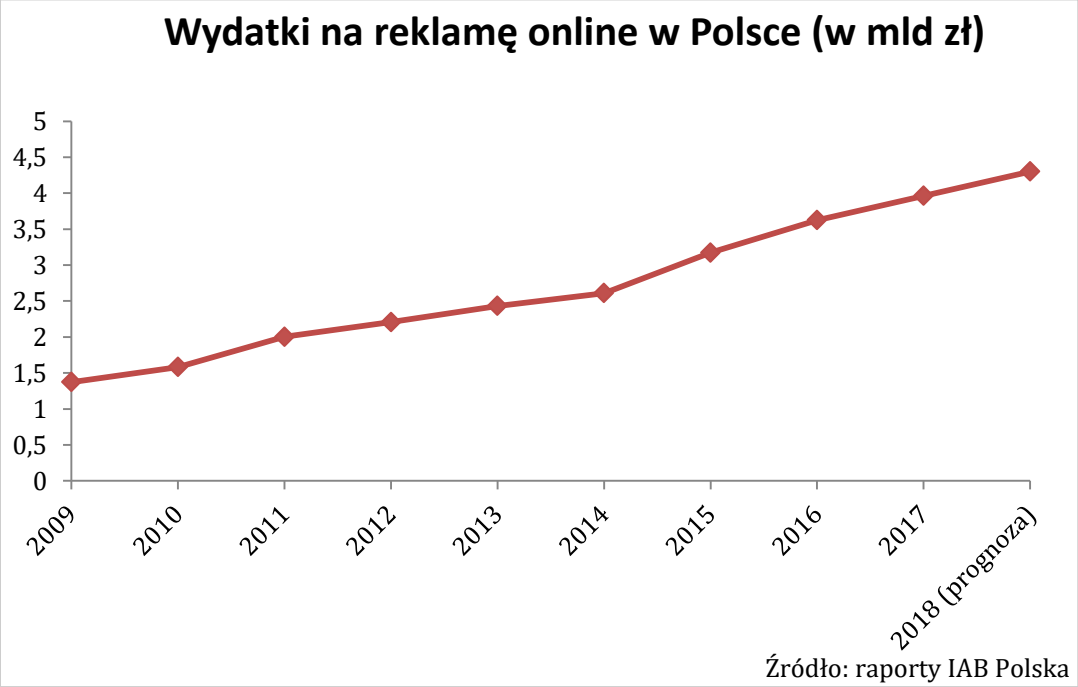
Analiza zainteresowań, potrzeb i słabości użytkowników Internetu służy w praktyce do ich kategoryzowania i hierarchizowania. Ta wiedza – czy też raczej przeświadczenia – na temat (potencjalnych) klientów są następnie wykorzystywane do ich pozyskiwania i obsługiwania. Firmy opracowały strategie segmentacji klientów (segmentacja ułatwia dopasowanie klienta do konkretnej kampanii reklamowej), taktyki wpływania na ich zachowania oraz metody mierzenia i optymalizowania wyników takich działań. Strategie i modele biznesowe firm w coraz większym stopniu opierają się na ilości i jakości danych osobowych, jakie są w stanie pozyskać i skomercjalizować.

> **Bezpośrednią ofiarą tej pogoni za danymi jest prywatność użytkowników cyfrowych technologii, nieświadomych toczonej za ich plecami gry. Najdrobniejsze i najpoważniejsze decyzje zakupowe, codzienne zachowania**

[2] Komisja Europejska, *European Data Market Study*, https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy.

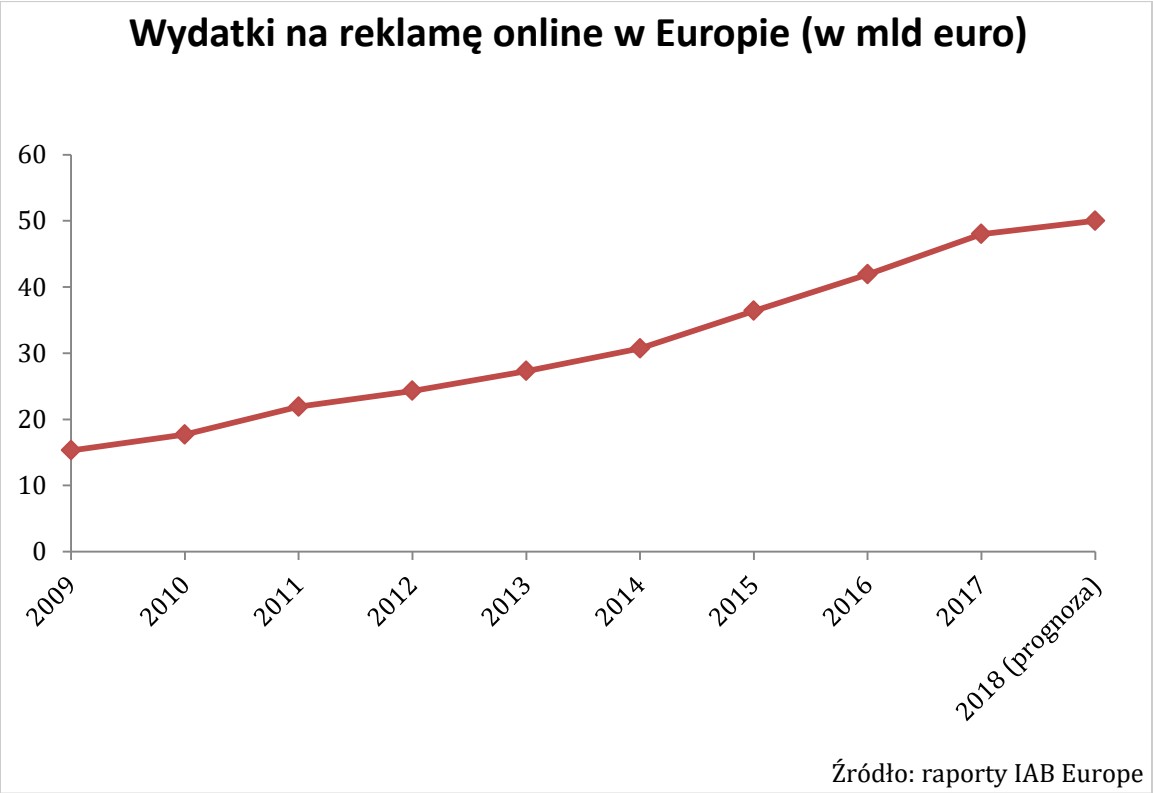
w sieci (od pojedynczego ruchu myszką i kliknięcia po zalogowanie na stronie internetowej), schematy przemieszczania się, relacje społeczne, zainteresowania oraz najbardziej intymne chwile miliardów ludzi są na bieżąco rejestrowane, analizowane i oceniane.

W tym kontekście trudno się dziwić, że sektor obsługujący przetwarzanie danych osobowych na potrzeby marketingu osiąga zawrotne zyski. Tylko w Polsce prognozowane wydatki na reklamę online w 2018 r. wyniosły 4,3 mld złotych[3]. Na poziomie europejskim w ciągu pięciu lat wartość rynku wzrosła dwukrotnie, osiągając w 2017 r. wysokość 48 mld euro. Te szacunki dotyczą wyłącznie wydatków na powierzchnię reklamową. Nie są nimi objęte m.in. koszty produkcji czy koszty zakupu danych. Z tego powodu rzeczywista wartość rynku reklamy internetowej może być nawet wielokrotnie większa.



Źródło: raporty IAB Polska

[3] Reklama online obejmuje wszystkie formaty reklamy cyfrowej: reklamę *display* (reklama graficzna, np. banery na stronie, sponsorowane treści, wyskakujące okienka, akcje konkursowe), e-mail marketing, ogłoszenia i SEM (tekstowe reklamy w wyszukiwarkach i reklamy kontekstowe).



1. Urządzenia mobilne, aplikacje i sensory

Katalizatorem i kluczowym narzędziem śledzenia i profilowania na potrzeby komercyjne stały się smartfony – komputery osobiste nowej generacji, z którymi większość użytkowników nie rozstaje się przez całą dobę, w jednym urządzeniu łącząc sferę osobistą i zawodową oraz angażując się w setki (o ile nie tysiące) drobnych interakcji dziennie[4]. Standardowo korzystanie z aplikacji mobilnych generuje metadane: adres IP, czas dostępu, długość trwania sesji, rodzaj używanego oprogramowania, lokalizację urządzenia[5]. Z metadanych – w połączeniu z informacjami o tym, w jaki sposób określona osoba korzystała z aplikacji czy usługi (gdzie kliknęła, czego szukała, co kupiła, jak szybko pisała) – powstaje dokładny profil użytkownika, obejmujący także jego cechy osobowości oraz opis przyzwyczajeń i indywidualnego trybu życia. Takie dane mają niejednokrotnie wrażliwy charakter i głęboko ingerują w prywatność: wystarczy wspomnieć o aplikacjach

[4] Badania pokazują, że średnio dotykamy telefonu ponad 2000 razy dziennie, por. np. https://businessinsider.com.pl/technologie/nauka/uzaleznienie-od-telefonu-badacze-policzyli-ile-razy-go-dotykamy/x45yftb.

[5] Me and My Shadow, *Location Tracking*, https://myshadow.org/location-tracking.

typu *fitness tracker* czy zdobywających coraz większą popularność aplikacjach rejestrujących cykl menstruacyjny i seksualne zachowania kobiet.

Przeciętny użytkownik smartfona korzysta aktywnie z około 27 aplikacji miesięcznie. Tylko nieliczni inwestują czas i uwagę w zapoznanie się z ich regulaminami i politykami prywatności. Badanie przeprowadzone przez naukowców z Carnegie Mellon University wykazało, że przeciętna polityka prywatności liczy aż 2518 słów, a jej przeczytanie zajmuje ok. 10 minut[6]. Autorzy raportu *Appfail Report* wskazują, że polityki prywatności najpopularniejszych aplikacji są nie tylko długie, ale też często nieczytelne i niezrozumiałe dla potencjalnego użytkownika[7].

Dark patterns

Dark patterns (w wolnym tłumaczeniu: „wredne praktyki") to elementy interfejsów w aplikacjach i serwisach internetowych zaprojektowane tak, by skłonić użytkownika do wybrania opcji najkorzystniejszej dla firmy zarabiającej na komercjalizacji danych. Kolory, rozmieszczenie przycisków i sposób wyświetlania komunikatów mają doprowadzić do tego, że użytkownik (niekoniecznie świadomie) „zgodzi się" na głęboką ingerencję w prywatność (np. ciągłe śledzenie lokalizacji czy dopuszczenie skryptów śledzących obsługiwanych przez podmioty trzecie).

Norweska Rada ds. Konsumentów w wydanym w 2018 r. raporcie *Deceived by Design* przeanalizowała, jak Facebook, Google i Microsoft (a dokładniej Windows 10) zaprojektowały interfejsy ustawień prywatności[8]. Okazało się, że każda z tych firm stosuje różnego rodzaju wredne praktyki, które mają skłonić użytkowników do oddania większej ilości danych, niż to konieczne do świadczenia usługi.

Najczęściej stosowane *dark patterns* to:

- niekorzystne dla użytkowników ustawienia domyślne;
- wydłużona ścieżka zmiany ustawień dla tych, którzy chcą zadbać o swoją prywatność;

---

[6] Por. raport A. M. McDonald, L. F. Cranor, *The Cost of Reading Privacy Policies*, http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf.

[7] The Norwegian Consumer Council, *Appfail Report – Threats to Consumers in Mobile Apps*, https://www.forbrukerradet.no/undersokelse/2015/appfail-threats-to-consumers-in-mobile-apps/.

[8] The Norwegian Consumer Council, *Deceived by Design. How tech companies use dark patterns to discourage us from exercising our right to privacy*, https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf.

- wyskakujące okienka z ustawieniami prywatności, w których kluczowe informacje zostały pominięte lub przedstawione w mylący sposób;
- grożenie utratą ważnych funkcjonalności lub usunięciem konta, jeśli użytkownik nie „zgodzi się" na przekazanie dodatkowych danych;
- interpretowanie nierzadko przypadkowych akcji (np. nieznacznego ruchu myszką) na korzyść aplikacji/serwisu.

W efekcie większość użytkowników aplikacji i urządzeń mobilnych rozpoczyna korzystanie z nich bez świadomości podstawowych funkcji i procesów związanych z przetwarzaniem danych osobowych. Wielu z nich zapewne nigdy nie wyraziłoby zgody na bardziej ryzykowne transakcje (np. przekazanie ich danych innym podmiotom bez związku z istotą usługi czy permanentne lokalizowanie), ale wychodzą z założenia, że nie mają wyboru[9]. Do głębszej analizy polityk prywatności zniechęca brak wpływu użytkowników na ich kształt: jedyne, co mogą zrobić, to pójść gdzie indziej. Często tylko po to, by się przekonać, że czeka na nich równie zły regulamin i pełna wytrychów prawnych polityka prywatności, albo zderzyć się z faktycznym brakiem konkurencyjnych usług. Ostatecznie korzystanie z wielu urządzeń mobilnych nie jest dziś możliwe bez konta na serwerze Google'a, Microsoftu czy Apple'a.

Na co „zgadzają się" użytkownicy?

Wystarczy pobieżnie przejrzeć polityki prywatności najpopularniejszych aplikacji mobilnych, aby odtworzyć rynkowy „standard przetwarzania danych osobowych", z jakim mierzy się typowy użytkownik smartfona. I tak – bez względu na rodzaj i funkcje zainstalowanej aplikacji – ów użytkownik zapewne będzie musiał zezwolić jej na:

- dostęp do swoich kontaktów,
- dostęp do kalendarza,
- dostęp do historii przeglądanych stron i zakładek,
- dostęp do wrażliwych logów aplikacji systemowych,
- dostęp do aplikacji aktywnych na danym urządzeniu,

[9] Por. J. A. Obar, A. Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465.

- dostęp do historii wybieranych numerów,
- dostęp do wszystkich profili użytkownika na danym urządzeniu,
- dostęp do treści i metadanych wysyłanych SMS-ów,
- dostęp do załączników e-maili,
- możliwość zmiany ogólnych ustawień telefonu.

Nie wszystkie z powyższych upoważnień i nie w każdym kontekście muszą być problematyczne. Na przykład uprawnienie do dostępu do aktywnych aplikacji zapewne nie wzbudzi zastrzeżeń w przypadku aplikacji służących do zarządzania zadaniami (często zintegrowanych z kalendarzem czy kontem pocztowym). Jeśli jednak o taki dostęp „prosi" oprogramowanie typu latarka czy *fitness tracker*, jest to już powód do niepokoju. Aktywność użytkownika w ramach innych aplikacji to przecież bogate źródło wiedzy o jego preferencjach i zwyczajach.

Tę samą logikę można odnieść do udostępniania kontaktów zgromadzonych w telefonie – do czego jest on potrzebny aplikacji, która nie oferuje żadnej funkcji związanej z zarządzaniem kontaktami? I tak dalej. Oczywiście, natura tych pytań jest retoryczna. Twórcy aplikacji nie kryją, że często ich główną intencją i powodem, dla którego oferują coś za darmo, jest zebranie danych, których sama aplikacja nie wykorzystuje, ale które mają realną rynkową wartość. Dlatego większość aplikacji dostępnych na rynku przekazuje dane stronom trzecim.

Aplikacje, które nie są tym, czym się wydają

Kategoria 1:	Złośliwe lub fałszywe aplikacje
Przykład	BZWBKlight, producent: West Corp Services
Co o sobie mówi?	Lżejsza wersja aplikacji mobilnej banku BZ WBK
Co tak naprawdę robi?	Podszywa się pod aplikację banku i **przechwytuje SMS-y**.

Kategoria 2:	Aplikacje wykradające dane na potrzeby reklamy interaktywnej
Przykład	250 gier mobilnych firmy Aphonso, np. *Honey Quest*, *Pool 3D*, *Endless 9*9 Puzzle*

Co o sobie mówi?	Gra mobilna
Co tak naprawdę robi?	Prosi o dostęp do mikrofonu i lokalizacji. Dopiero po wejściu w ustawienia wyjaśnia, że **śledzi nawyki oglądania telewizji w celu serwowania reklam**.

Kategoria 3: Serwisy społecznościowe

Przykład	Facebook
Co o sobie mówi?	Serwis społecznościowy, który „zbliża ludzi z całego świata do siebie" (z ang. *bring the world closer together*)
Co tak naprawdę robi?	Ostatnie wnioski patentowe Facebooka: • **automatyczne uruchamianie mikrofonu** co jakiś czas, aby wykryć, jakie programy użytkownik ogląda w telewizji i czy np. wycisza telewizor na czas przerwy reklamowej; • system, który na podstawie zdjęć i ich opisów oraz wspólnych numerów IP i identyfikatorów urządzeń **pomaga ustalić, z kim użytkownicy dzielą gospodarstwo domowe**; • Offline Trajectories – technologia, która ma **przewidywać przyszłe lokalizacje użytkowników**. W obu rozwiązaniach chodzi o wydobywanie dodatkowych danych i **komercjalizowanie jak najpełniejszych profili użytkowników**.

Kategoria 4: Aplikacje do uczenia sztucznej inteligencji

Przykład	Google Arts & Culture, Art Selfie
Co o sobie mówi?	„Zabawny sposób na poznanie sztuki" – użytkownik robi sobie selfie, a sztuczna inteligencja Google'a porównuje zdjęcie do dzieł sztuki.
Co tak naprawdę robi?	Miliony wrzucanych przez użytkowników zdjęć Google wykorzystuje **do trenowania swojej sztucznej inteligencji**.

Skąd aplikacje wiedzą, gdzie jesteś?

Aplikacje mobilne nagminnie śledzą naszą lokalizację. Dzieje się tak przede wszystkim dlatego, że mają do niej łatwy dostęp, a pozyskana w ten sposób wiedza ma realną wartość na rynku reklamowym. Do ustalenia naszej (przybliżonej) lokalizacji aplikacje wykorzystują:

- dane o logowaniu telefonu do stacji przekaźnikowych;
- dostęp do czipu GPS zainstalowanego w telefonie lub do wewnętrznego rejestru lokalizacji (*location logs*);
- historię zapamiętanych sieci Wi-Fi (nawet jeśli nie doszło do połączenia z siecią);
- historię przypisanych naszemu urządzeniu adresów IP;
- metadane zapisanych na naszym urządzeniu zdjęć, które standardowo zawierają współrzędne geograficzne ustalone w momencie robienia zdjęcia (tzw. dane EXIF);
- tzw. odcisk palca przeglądarki (z ang. *browser fingerprint*);
- adres zapamiętany w przeglądarce lub aplikacji (np. „Dom" na Google Maps lub „Praca" w Uberze).

Ustalenie lokalizacji w oparciu o takie dane nie wymaga naszej wiedzy ani zgody. W praktyce dość trudno jest to wykryć, a jeszcze trudniej skutecznie wyłączyć. Nasze pomysły na ochronę prywatności zwykle się kończą na wyłączeniu funkcji lokalizacji GPS w telefonie. Twórcy aplikacji doskonale to wiedzą i dlatego tak chętnie zastrzegają sobie dostęp do mniej oczywistych wskaźników.

Nowa jakość śledzenia: wszechobecne sensory

Techniki mobilnego śledzenia mimo wysokiej skuteczności stale ewoluują. W ostatnich latach weszły na kolejny poziom dzięki możliwości wykorzystania sensorów w rozmaitych urządzeniach, które wkroczyły do biur i mieszkań. Czytniki e-booków, smart TV, termostaty, czujniki gazu, inteligentne lodówki, okulary, szczoteczki do zębów, zabawki i autonomiczne odkurzacze zasilają bazy danych nowymi wskaźnikami, zbieranymi i nierzadko przekazywanymi dalej w czasie rzeczywistym. Tego typu urządzenia – analogicznie do smartfonów – zapewniają firmom stały dostęp do informacji na temat zainteresowań i nawyków klientów.

Zbierane dane mogą być kontrolowane przez jedną firmę (tak jest np. w przypadku czytnika Kindle), jak również udostępniane stronom trzecim, najczęściej w celach marketingowych (w tym w celu zaproponowania kolejnych, także śledzących, aplikacji). Do baz danych, puchnących dzięki wszechobecnym sensorom, dostęp mogą mieć również tzw. platformy IoT – zaawansowane aplikacje ułatwiające firmom analizowanie i zarządzanie danymi, dostarczane i obsługiwane przez wyspecjalizowane korporacje.

W ekosystemie opartym na komercjalizacji danych osobowych nic się nie marnuje, a zysk może czerpać wiele niezależnych podmiotów. Dobrym przykładem jest wykorzystanie sensorów i inteligentnego oprogramowania w samochodach. Terabajty danych generowane przez inteligentne auta to realna wartość dla producentów, dealerów samochodowych, leasingodawców, autorów wyspecjalizowanych w tej branży aplikacji, producentów oprogramowania do telefonów komórkowych (takich jak Google czy Apple), brokerów danych, agencji reklamowych, jak również służb porządkowych czy windykatorów – a lista jest znacznie dłuższa. Każdy z tych profesjonalnych graczy jest w stanie znaleźć coś dla siebie, nie odbierając wartości innym.

W najgorszej pozycji jest sam użytkownik, którego żaden z wymienionych powyżej podmiotów nie ma ochoty informować o swoich interesach ani tym bardziej prosić o zgodę na skorzystanie z tak łatwo dostępnego bogactwa. Jeśli ten trend nie zostanie odwrócony, systemy połączonych urządzeń działające w ramach tzw. Internetu rzeczy (ang. IoT – *Internet of Things*) mogą stać się najpoważniejszym, bezprecedensowym zagrożeniem dla prywatności, a w niektórych przypadkach (np. inteligentnych samochodów czy sprzętu medycznego) także fizycznego bezpieczeństwa ich użytkowników.

Smart dom czy permanentna inwigilacja? Przegląd wybranych urządzeń

Urządzenie:	Roomba – smart odkurzacz
Co robi?	Dzięki czujnikom odtwarza plan mieszkania, po którym się porusza – jego powierzchnię oraz odległości pomiędzy meblami i urządzeniami.
Problemy	iRobot – producent Roomby – zapowiedział, że firma rozważa udostępnienie planów mieszkań Amazonowi, Apple'owi i Google'owi – internetowym gigantom coraz bardziej aktywnym w branży smart urządzeń. Dzięki danym z Roomby firmy te mogłyby łączyć z odkurzaczem swoje urządzenia i polecać użytkownikom inne produkty.

Urządzenie:	Fredi – monitor dziecięcy
Co robi?	Transmituje obraz i dźwięk z pomieszczenia, w którym przebywa dziecko.
Problemy	Wielokrotnie występowały przypadki wycieków danych i ataków hakerskich; słabe hasło, brak szyfrowania danych, brak automatycznych aktualizacji oprogramowania.

Urządzenie:	Pompa powietrzna CPAP
Co robi?	Wspiera oddychanie podczas snu u osób cierpiących na zespół bezdechu sennego.
Problemy	Pompa wysyła informacje o tym, kiedy i jak długo użytkownik z niej korzysta, nie tylko do lekarza, ale też do producenta urządzenia oraz do ubezpieczyciela. Jeśli dane pokazują, że pompa nie jest regularnie wykorzystywana, ubezpieczyciel może odmówić jej finansowania.

Urządzenie:	Amazon Echo
Co robi?	Sterowany głosem smart głośnik, przez który komunikuje się Alexa – asystentka głosowa Amazona.
Problemy	Amazon zapewnia, że głośnik zaczyna słuchać dopiero wtedy, gdy użytkownik wypowie słowo „Alexa". Tymczasem w 2018 r. Amazon Echo nagrał prywatną rozmowę swojej użytkowniczki i bez jej zgody wysłał nagranie przypadkowej osobie na liście kontaktów. Amazon tłumaczył, że było to spowodowane błędem systemu. Tymczasem firma zarejestrowała wniosek patentowy na system, który słucha rozmów użytkowników i wyłapuje opinie o produktach, żeby lepiej personalizować reklamę.

Urządzenie:	Vizio Smart TV
Co robi?	Smart telewizor
Problemy	Domyślne ustawienia pozwalały telewizorowi zbierać dane, sekunda po sekundzie, o wszystkim, co użytkownik oglądał. Dane były sprzedawane reklamodawcom, którzy

dzięki adresowi IP mogli je integrować ze swoimi informacjami o tych samych osobach.
Po interwencji ze strony amerykańskiej Federalnej Komisji Handlu Vizio zaprzestał tych praktyk.

2. Skrypty śledzące i targetowane reklamy

Dostawcy treści w Internecie (wydawcy gazet, właściciele blogów, media społecznościowe) utrzymują się przede wszystkim ze sprzedaży przestrzeni reklamowej. Ten model finansowania oznacza, że ich realnym klientem jest nie ten, kto odwiedza stronę, ale ten, kto płaci za wyświetlenie na niej reklamy.

Marketingowa przewaga mediów internetowych nad tradycyjną prasą i telewizją wzięła się z nieporównanie lepszych możliwości targetowania przekazu. Czasy, w których reklamodawcy inwestowali w kampanie kierowane do „kobiet z dużych miast" czy „mężczyzn w pewnym przedziale wiekowym", są daleko za nami. Wygrywa reklama targetowana, skierowana do zdefiniowanego odbiorcy – i to w momencie, w którym jest szansa, że ten konkretny człowiek kupi oferowany mu produkt.

Reklama interaktywna, czyli co?

Reklama internetowa – każdy format reklamowy wykorzystywany w Internecie. Najpopularniejsze są reklamy typu *display* (np. graficzne banery), reklamy w wyszukiwarkach oraz e-mail marketing.

Reklama kontekstowa – reklama dopasowana do treści strony, na której się wyświetla (np. w artykule o samochodach użytkownik zobaczy reklamy konkretnych marek).

Reklama targetowana / behawioralna / profilowana – trzy określenia na ten sam typ reklamy: dopasowanej nie do treści strony, ale do preferencji użytkownika, który ją odwiedza.

W zależności od swoich relacji biznesowych i posiadanej ilości danych o użytkownikach dostawcy sprzedają przestrzeń bezpośrednio domom mediowym reprezentującym reklamodawców albo korzystają z usług pośredników, wystawiając dostępną przestrzeń

i profil użytkownika na giełdzie reklam (tzw. **model RTB**, z ang. *real-time bidding*). O ile sprzedaż bezpośrednią wydawca może prowadzić albo rękoma swoich pracowników, albo w sposób zautomatyzowany (z wykorzystaniem tzw. technologii *programmatic*), o tyle sprzedaż za pośrednictwem giełdy jest możliwa tylko w modelu programatycznym.

Ostatnie lata przyniosły **zmianę oczekiwań reklamodawców, dla których liczy się już nie to, do ilu osób dotarł ich przekaz, ale ile z nich dobrze na reklamę zareagowało – a więc jej efektywność**. W konsekwencji zarobki wydawców coraz częściej są uzależnione nie od liczby wyświetleń ani nawet kliknięć w reklamę, ale od liczby osób, które zostawiły reklamodawcy swoje dane kontaktowe (tzw. model rozliczeniowy CPL – *cost per lead*) lub faktycznie kupiły dany produkt (tzw. model CPS – *cost per sale*).

W związku z tą zmianą rynku reklamowego koszty bezpośredniej sprzedaży powierzchni po stronie wydawców rosną. Jeśli chcą zagwarantować reklamodawcom realne efekty, potrzebują coraz więcej danych o użytkownikach i coraz lepszych działów reklamy. Z tego powodu wydawcy ograniczyli liczbę powierzchni dostępnych w sprzedaży bezpośredniej do najważniejszych klientów i zwiększyli udział reklam sprzedawanych w modelu RTB. W tym modelu dopasowanie reklamy do użytkownika jest wspomagane danymi z różnych źródeł, a więc nie polega tylko na działaniach wydawcy.

Rosnący popyt na jak najlepiej stargetowaną reklamę zwiększa atrakcyjność śledzenia użytkowników. Tylko znając ich zainteresowania, siłę nabywczą i wzorce behawioralne, dostawcy treści są w stanie zapewnić swoim klientom odpowiedni wskaźnik konwersji i uzyskać dobrą cenę za wyświetlane reklamy. Taka operacja marketingowa wymaga rzetelnych danych na temat wieku, płci, miejsca zamieszkania, zainteresowań, sytuacji życiowej, sytuacji majątkowej, historii zakupów, sieci społecznościowej, a nawet cech osobowości[10] potencjalnego klienta. Ilu z nich ujawnia je świadomie, z własnej woli? Niewielu. Nawet świadomość tego, że tak szczegółowe dane są zbierane lub generowane (na zasadzie predykcji) w celach marketingowych, nie jest powszechna[11].

Profile marketingowe stanowią zbiór wielu pozornie nieistotnych okruchów danych, które nieświadomi użytkownicy zostawiają za sobą w sieci: numerów IP, logów, informacji

[10] https://en.wikipedia.org/wiki/Big_Five_personality_traits.

[11] Por. badania The Chartered Institute of Marketing, *Whose data is it anyway?*, https://exchange.cim.co.uk/blog/consumers-in-the-dark-about-their-own-data/.

„Some 92% of respondents did not fully understand how information that companies gleaned about them was being used, and they were highly sceptical about marketing practices" oraz raport: T. Morey, A. Schoop, *Customer Data: Designing For Transparency And Trust*, potwierdzający, że aż 69% respondentów jest zaniepokojnych samą informacją, że ich dane mogłyby być wykorzystywane do innych celów niż zadeklarowane (w tym marketingowych). Według Eurobarometru aż 71% badanych uważa, że niedopuszczalnym jest, by firmy dzieliły się danymi o nich, nawet gdyby miało to poprawić jakość usług, z których lubią korzystać, https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust.

o lokalizacji, pojedynczych kliknięć, drobnych transakcji online, historii odwiedzanych stron i pytań zadawanych wyszukiwarce. W jaki sposób i przez kogo są wykorzystywane? Rzetelna odpowiedź na to pytanie nie jest prosta. Wymaga zagłębienia się w dyskretny, działający na „zapleczu" Internetu ekosystem reklamy behawioralnej. W żargonie marketingowym określa się go poprzez kluczową funkcję, jaką pełni: *real-time bidding* (RTB) – w wolnym tłumaczeniu: „licytacje w czasie rzeczywistym". Przedmiotem owych licytacji są tzw. impresje (z ang. *impression*), czyli pojedyncze wyświetlenia reklam przez potencjalnych odbiorców – użytkowników o określonym profilu, którzy w danym momencie przeglądają stronę.

Jak działa mechanizm *real-time bidding* (RTB)?

Dostawcy treści internetowych wystawiają na sprzedaż profile swoich użytkowników oraz wyświetlaną im przestrzeń reklamową za pośrednictwem tzw. **platform podaży** (ang. *Supply Side Platforms*, **SSP**) – wyspecjalizowanego oprogramowania, które w czasie rzeczywistym umożliwia komunikację z innymi graczami na giełdzie reklam.

Oferowany profil nie zawiera danych bezpośrednio identyfikujących użytkownika, takich jak nazwisko czy adres. Te informacje z perspektywy reklamodawców mają niewielką wartość. Liczą się tylko cechy decydujące o skłonności użytkownika do dokonania konkretnego zakupu. Dlatego pośrednicy posługują się unikatowymi numerami użytkownika, do których przypisany jest profil o znaczeniu marketingowym (zainteresowania, siła nabywcza, stan zdrowia, ważny moment w życiu etc.).

Użytkownicy korzystający ze smartfonów i innych urządzeń mobilnych są znani różnym pośrednikom pod jednym, stałym identyfikatorem reklamowym – na urządzeniach z Androidem jest to Google Advertising ID (AAID), z kolei urządzenia Apple korzystają z numeru IDFA (ang. *Identifier for Advertising*). W świecie laptopów i komputerów stacjonarnych jest inaczej – ilu dostawców treści i pośredników, tyle różnych identyfikatorów. Poszczególnym graczom w zorientowaniu się, kto jest kim, pomaga tzw. ***cookie syncing*** (dopasowywanie/synchronizowanie ciasteczek) – proces, w ramach którego firmy wzajemnie przekazują sobie informacje o nadanych użytkownikom identyfikatorach i kryjących się pod nimi profilach. Katalogowaniem identyfikatorów i profili od różnych pośredników zajmują się wyspecjalizowane **platformy zarządzania danymi** (ang. *Data Management Platforms*, **DMP**).

Dane wystawiane przez sprzedawców przestrzeni reklamowej odbierają i analizują tzw. **platformy popytu** (ang. *Demand Side Platforms*, **DSP**), zaprogramowane tak, by wyszukiwać użytkowników o określonym profilu. Profile poszukiwanych użytkowników określają **agencje mediowe** – kluczowi rozgrywający po tej stronie rynku reklamowego.

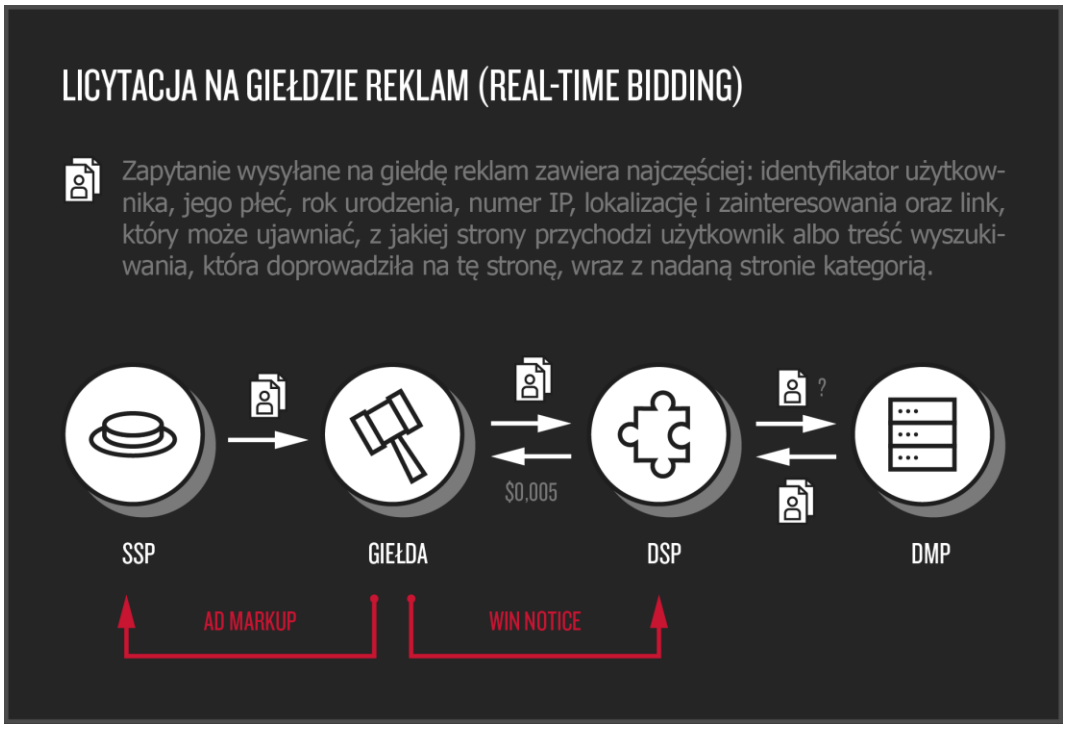
Ich klienci przychodzą ze standardowym zleceniem: „Daj mi klienta, który kupi po jak najlepszej cenie to, co mam do sprzedania".

Zadaniem agencji jest określenie, jakie cechy ma taka osoba i gdzie ją można znaleźć w sieci. Podstawą ich wiedzy o potencjalnych klientach są dane o zakupach pochodzące bezpośrednio od reklamodawców. Aby je wygenerować, większość firm utrzymuje systemy zarządzania relacjami (CRM), programy lojalnościowe i własne sklepy internetowe. Z informacji o tym, kto i co kupił w przeszłości, wyciągane są cechy typowego klienta. To punkt wyjściowy dla agencji mediowych, które mierzą dalej: próbują ustalić, jakie cechy i zachowania charakteryzują osoby, które jeszcze reklamowanego produktu nie kupiły, ale mogą to zrobić w przyszłości. W żargonie reklamowym ten hipotetyczny profil klienta to *look-alike*.

> ***Look-alike* to mieszanka twardych danych pochodzących od reklamodawców („taki człowiek już u mnie kupił") i statystycznej wiedzy o ludziach („taki człowiek może chcieć to kupić"). Komponent statystyczny pochodzi z analizy danych zebranych przez dostawców treści, media społecznościowe i brokerów danych i wystawionych na sprzedaż agencjom mediowym**.

Zadaniem **giełd reklamowych** (ang. *ad exchanges*) jest optymalnie dopasować reklamę do użytkownika, który powinien ją zobaczyć. Po stronie podaży pojawia się informacja o profilu konkretnego użytkownika (który w tym momencie czeka na załadowanie strony internetowej, a więc można mu wyświetlić reklamę), po stronie popytu „czeka" profil *look-alike*, którego poszukują agencje mediowe. Na ich dopasowanie giełdy mają ułamki sekund.

Standardowa transakcja w modelu RTB rozpoczyna się w momencie nawiązania przez przeglądarkę połączenia ze stroną, którą ktoś właśnie próbuje załadować na swoje urządzenie. Strona ustala profil tego użytkownika, przypisuje mu unikatowy numer lub sprawdza, czy ten użytkownik nie figuruje już w jej bazie, i powiadamia **serwer reklam**. Serwer reklam łączy się z **platformą podaży (SSP)** i przekazuje jej dane o dostępnej przestrzeni reklamowej oraz informacje o użytkowniku, który czeka na załadowanie reklamy, takie jak: nadany mu identyfikator; link, który może ujawniać, z jakiej strony przychodzi użytkownik albo jak brzmiało jego wyszukiwanie, które doprowadziło na tę stronę; rok urodzenia, płeć czy zainteresowania (zob. więcej w ramce poniżej). Platforma podaży wystawia ten profil na giełdzie i czeka na ofertę.



Giełda reklam rozsyła unikatowy numer sprofilowanego użytkownika do współpracujących z nią platform popytu (DSP). Ten moment w żargonie reklamowym to ***bid request***. Potencjalni oferenci dowiadują się, że osoba o danym identyfikatorze właśnie czeka na wyświetlenie strony internetowej i że mogą dołączyć do niej swoją reklamę. Czy się na to zdecydują i ile zapłacą, zależy od tego, w jakim stopniu profil kryjący się pod identyfikatorem pokrywa się z poszukiwanym przez nich *look-alike*.

Jakie dane wysyłane są na giełdę reklam?

Proces licytacji w modelu RTB jest ustandaryzowany – wszystkie firmy, które chcą wziąć udział w licytacji, muszą dostosować się do szczegółowych wytycznych technicznych autorstwa Google'a i Internet Advertising Bureau (branżowej organizacji firm działających na rynku reklamy interaktywnej). Te dwa podmioty stworzyły dominujące systemy RTB (IAB – OpenRTB, Google – Authorized Buyers) i to one ustalają standardy techniczne i organizacyjne dla pozostałych graczy. Wytyczne dyktują m.in. to, jakie informacje o użytkowniku są wysyłane przez platformę podaży (SSP) na giełdę, a w konsekwencji – jakie dane trafią do setek licytujących platform popytu (DSP).

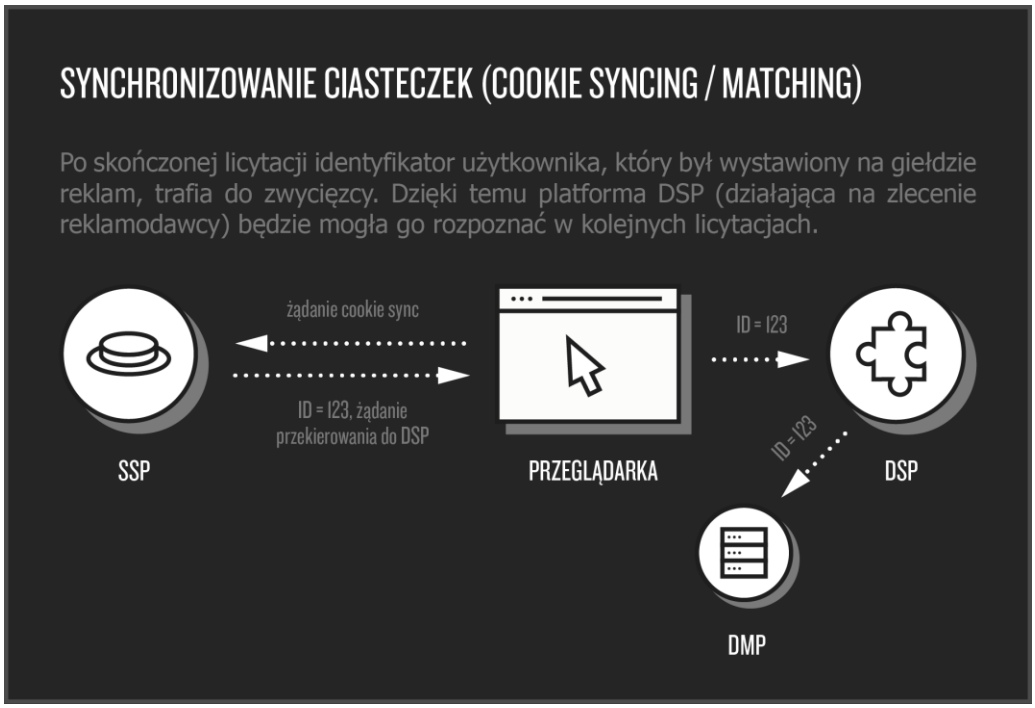
Standardowe zapytanie wysyłane na giełdę (*bid request*) zawiera: 1) identyfikator nadany użytkownikowi przez SSP, 2) tzw. *full referral URL*, czyli link do strony, na której ma się wyświetlić reklama, treść zapytania zadanego wyszukiwarce lub adres strony, z której przychodzi użytkownik, oraz nadaną stronie kategorię (np. pomoc ofiarom przemocy), a ponadto – jeśli SSP ma takie informacje: 3) rok urodzenia, 4) płeć, 5) lokalizację, 6) numer IP, 7) zainteresowania, a także 8) inne dane, które firma ma na temat użytkownika. **Wszystkie te informacje mogą ujawniać dane wrażliwe.**

Attribute	Type	Definition
`id`	string; recommended	Vendor-specific ID for the user. At least one of `id` or `buyeruid` is strongly recommended.
`buyeruid`	string; recommended	Buyer-specific ID for the user as mapped by an exchange for the buyer. At least one of `id` or `buyeruid` is strongly recommended.
`yob`	integer	Year of birth as a 4-digit integer.
`gender`	string	Gender, where "M" = male, "F" = female, "O" = known to be other (i.e., omitted is unknown).
`keywords`	string	Comma separated list of keywords, interests, or intent.
`consent`	string	GDPR consent string if applicable, complying with the comply with the IAB standard Consent String Format in the Transparency and Consent Framework technical specifications.
`geo`	object	Location of the user's home base (i.e., not necessarily their current location). Refer to Object: Geo.
`data`	object array	Additional user data. Each `Data` object represents a different data source. Refer to Object: Data.
`ext`	object	Optional vendor-specific extensions.

Obrazek przedstawia niektóre informacje zawarte w *bid request*. Źródło: IAB[12].

[12] https://github.com/InteractiveAdvertisingBureau/AdCOM/blob/master/AdCOM%20v1.0%20FINAL.md.

W ułamkach sekund, które mają na zawarcie transakcji, platformy popytu sprawdzają, czy użytkownik o tym identyfikatorze znajduje się w ich bazie, oraz poprzez platformy DMP zaciągają dodatkowe dane na temat wystawionego profilu. Co wiadomo o jego przeszłych transakcjach? Jakie strony odwiedził przed chwilą? Czy porównywał ceny? Czego szuka i ile jest w stanie zapłacić? Korzystając ze skryptów śledzących, a czasem też dodatkowych informacji pozyskanych od brokerów danych, podejmują decyzję (tak/nie) i wysyłają ofertę. Oferent, który zaproponował najwyższą stawkę, wygrywa prawo do wyświetlenia reklamy i zapisania danych identyfikujących użytkownika w swojej bazie.



Proces licytacji jest w pełni zautomatyzowany i zdominowany przez algorytmy. Z punktu widzenia profilowanego użytkownika pozostaje całkowicie niezauważalny – cała transakcja zajmuje 1/5 sekundy (200 milisekund). To szybciej niż mrugnięcie okiem, na które przeciętnie potrzebujemy „aż" 300 milisekund. Dlatego tylko nieliczni interesują się tym, co się dzieje na komercyjnym „zapleczu" Internetu. W rezultacie mamy do czynienia z niebezpieczną asymetrią informacyjną, którą profesjonalni gracze wykorzystują, by – w nieujawniony sposób – wpływać na zachowania nieświadomych użytkowników.

Ile kosztują Twoje dane i kto ma do nich dostęp?

Pojedyncze wyświetlenie (impresja) jest warte średnio 0,0005 dolara, czyli około 1/5 grosza. Na jej wartość wpływa to, jak bogaty jest profil użytkownika i na ile pasuje do cech, których aktualnie poszukuje reklamodawca. Łukasz Olejnik, ekspert ds. bezpieczeństwa i prywatności, ustalił, że cena może również zależeć od lokalizacji użytkownika (profile internautów z USA są dwa razy droższe niż tych z Francji), a nawet od pory dnia – np. ceny impresji są wyższe między północą a 8 rano, ponieważ mniejsze grono przeglądających strony internetowe użytkowników oznacza większą konkurencję na giełdzie.

Co się dalej dzieje z danymi osobowymi, które pojawiły się w zaproszeniu do składania ofert? W modelu RTB identyfikator użytkownika i podstawowe informacje na jego temat mogą trafić do setek pośredników. Licytacja odbywa się zazwyczaj nie na jednej, ale na kilku giełdach reklam jednocześnie, przy czym na każdej z nich o zdobycie „impresji" walczą w imieniu reklamodawców dziesiątki, może nawet setki platform. **Wszystkie te podmioty mają dostęp do profilu użytkownika przesłanego na giełdę.**

Regulaminy największych giełd reklamowych, np. Authorized Buyers należącej do Google'a, zastrzegają, że tylko podmiot, który wygrał licytację, może zatrzymać dane o użytkownikach w celu wzbogacenia swojej bazy. Inne firmy nie mogą agregować ich z informacjami pozyskanymi z innych źródeł ani profilować użytkowników na tej podstawie. Aby zdobyć informacje o użytkowniku, wystarczy jednak wygrać choć jedną aukcję (czyli zaoferować wyższą niż przeciętną cenę). W następnych aukcjach firma będzie już dysponowała dużo bogatszym profilem. Zobowiązanie do usunięcia danych o użytkownikach ma jednak wyłącznie umowny charakter. Nie wiadomo, czy i w jaki sposób jest egzekwowane. Dominujące systemy RTB autorstwa Google'a i IAB zostały tak skonstruowane, że **w praktyce nad raz rozpowszechnionymi na giełdzie danymi nie ma jakiejkolwiek kontroli – ani ze strony firmy, która wysłała dane na giełdę, ani tym bardziej ze strony samego użytkownika**.

Kim są kluczowi gracze?

Szybkie i precyzyjne dopasowanie profilu użytkownika, który w danym momencie przegląda strony internetowe, do poszukiwanego przez reklamodawców (*look-alike*) jest łatwiejsze, jeśli obie strony rynku reklamowego mają dostęp do tych samych danych o ludziach i tych samych technik profilowania. Z tej prawidłowości wynika przewaga firm

z najlepiej rozwiniętą analityką i dostępem do największych baz danych, szczególnie danych behawioralnych możliwych do połączenia z konkretnym użytkownikiem za pomocą adresu e-mail lub innego stałego identyfikatora.

Dlatego niekwestionowanymi liderami w zautomatyzowanym handlu przestrzenią reklamową są największe platformy internetowe: **Facebook i Google**. Ich przewaga polega nie tylko na gigantycznych ilościach danych, jakie zebrały o ludziach, ale też na dostępie do technologii, która gwarantuje odpowiednią prędkość transakcji. Google stworzył rynek targetowanej reklamy, na którym kontroluje wszystkie kluczowe role: sprzedaje i kupuje przestrzeń reklamową oraz dostarcza dane i analizę danych po obu stronach transakcji. W tym momencie żaden inny gracz nie może z nim konkurować.

W drugim szeregu działają **brokerzy danych**, czyli pośrednicy w handlu danymi, którzy zdobywają dane o użytkownikach Internetu za pośrednictwem innych firm (banków, sklepów, ubezpieczycieli, innych stron internetowych etc.). Wspólną cechą brokerów danych jest to, że użytkownicy nie stykają się z nimi bezpośrednio, przez co wielu nie ma nawet świadomości istnienia tych firm, nie mówiąc o kontroli nad tym, jakie dane i w jakich celach są przez nie wykorzystywane. Najwięksi gracze w tej kategorii to firmy amerykańskie, takie jak Acxiom, Experian czy Datalogix (należący do Oracle). Lokalizacja głównych siedzib nie przeszkadza im zbierać na masową skalę danych o konsumentach na całym świecie, również w Europie. Wykorzystują do tego publiczne i komercyjne bazy danych oraz media społecznościowe.

Sam tylko Acxiom zebrał informacje o 700 milionach konsumentów, przy średniej liczbie danych na temat jednej osoby na poziomie 3000. Oracle z kolei chwali się, że współpracuje z 1500 dostawcami danych, którzy zasilają 2 mld profili użytkowników. Przy takiej skali i głębokości śledzenia ludzkich zachowań nawet nietypowy konsument staje się przewidywalny. W Polsce najważniejsze firmy świadczące tego typu usługi to Cloud Technologies i Netsprint. Ta pierwsza w 2017 r. zarobiła ponad 70 mln zł i – jak twierdzi – jest w stanie dostarczyć klientom z całego świata ponad 25 mld profili[13]. Druga w swoich materiałach reklamowych porównywała się do pudełka plasteliny, z której może ulepić dowolny segment użytkowników, wykorzystując „do zabawy" 92 mln ciasteczek[14].

[13] https://www.onaudience.com/dataexchange.
[14] Fragment pochodzi z reklamy zamieszczonej w raporcie *Data Driven Marketing w Polsce*.

„Zaufani partnerzy”

Brokerów danych i innych pośredników z branży *ad tech* (zajmującej się obsługą techniczną transakcji reklamowych) portale internetowe traktują jak „zaufanych partnerów”. To w ich imieniu wydawcy mediów elektronicznych zbierają zgodę swoich użytkowników na śledzenie, profilowanie i dopasowywanie reklam. Niektóre polskie portale „ufają” nawet 700 takim firmom. Czy za każdym razem sprawdzają, jakie dane i w jakim celu są zbierane przez „zaufanych partnerów”? Niekoniecznie.

IAB (organizacja zrzeszająca firmy z branży interaktywnej) zaproponowała wspólny dla całej branży standard wymiany informacji o preferencjach użytkowników[15]. Chodzi o tzw. *daisy bits*, czyli skompresowaną informację o udzielonych zgodach na przetwarzanie danych, z uwzględnieniem celów przetwarzania i z listą firm, którym ta osoba chce udostępniać swoje dane. Te informacje są przekazywane w plikach *cookie*. Na tej podstawie każdy portal może ustalić preferencje swoich użytkowników i je respektować (a więc w ramach danej sesji „wpuszczać” ciasteczka i skrypty śledzące tylko tych firm, które obejmuje udzielona zgoda).

W praktyce wygląda to dużo gorzej. Standard IAB nie mówi nic o jakości zbieranych zgód, a wręcz dopuszcza globalną zgodę na instalowanie ciasteczek i skryptów śledzących, bez względu na ich cel i źródło. **Jedno kliknięcie „Przejdź dalej” na wyskakującym okienku to dla firm korzystających ze standardu IAB wystarczająca podstawa, by śledzić takiego użytkownika na tysiącach współpracujących stron internetowych.**

Wyobraźmy sobie taki przykład: użytkownik zgadza się na śledzenie przez „zaufanych partnerów” na stronie apteki internetowej, a następnie wchodzi na portal kulinarny i przegląda przepisy na dania o obniżonej zawartości tłuszczu. Apteka nadal śledzi jego aktywność i na stronie portalu z przepisami wykupuje reklamę suplementów diety na odchudzanie. Użytkownik może nie skojarzyć, że te reklamy nie wyświetlają mu się przypadkiem – w końcu udzielił zgody tylko na stronie apteki.

A co, jeśli użytkownik potraktował swój wybór poważnie, zajrzał w ustawienia prywatności i świadomie nie zgodził się na śledzenie przez „zaufanych partnerów”? Nadal nie ma gwarancji, że ta decyzja będzie respektowana. Standard IAB nie wymusza na wydawcach i reklamodawcach blokowania skryptów śledzących nawet w przypadku,

[15] IAB Tech Lab, *GDPR Transparency and Consent Framework,* https://iabtechlab.com/standards/gdpr-transparency-and-consent-framework/.

kiedy użytkownik nie wyraził na nie zgody. To powoduje, że wiele firm uruchamia swoje skrypty automatycznie, nie czekając na wybór użytkownika lub zwyczajnie go ignorując. Z tych samych powodów większość portali nie respektuje sygnału Do Not Track[16], który użytkownik może ustawić w swojej przeglądarce. A przecież trudno o jaśniejszy i mocniejszy komunikat dla branży reklamowej: zostawcie mnie w spokoju!

Funkcjonujemy w Internecie podwójnych standardów: na każdym kroku jesteśmy zapraszani do decydowania o tym, komu chcemy przekazać swoje dane, ale biznes reklamowy działa po staremu, tak jakby wszystko było dozwolone. Tymczasem portale, które wpuszczają ciasteczka i skrypty śledzące „zaufanych partnerów", udają, że nie ponoszą za ich działanie żadnej odpowiedzialności. W końcu użytkownik „sam się zgodził".

3. Skala i głębokość śledzenia

Zakres danych, jakie są dostępne na temat konkretnego użytkownika (warto pamiętać, że w systemie RTB każdy ma unikatowy numer identyfikacyjny, nadany mu przez konkretnego pośrednika), w dużej mierze zależy od rodzaju usług, z których taka osoba korzysta. W szczególności od tego, czy korzysta z mediów społecznościowych lub innych platform (e-commerce, poczta elektroniczna), które wymagają zalogowania i podania prawdziwych danych, a zarazem umożliwiają ciągłe gromadzenie danych behawioralnych w połączeniu ze stabilnym identyfikatorem. Z perspektywy firm działających na rynku targetowanej reklamy to idealna sytuacja. Ale nawet wobec braku zweryfikowanych i łatwo dostępnych danych dostawcy treści internetowych (a zarazem sprzedawcy przestrzeni reklamowej) i współpracujący z nimi brokerzy danych próbują tworzyć wielowarstwowe profile użytkowników, na które standardowo składają się:

- adresy, kod pocztowy, płeć, wiek, stan cywilny, poziom edukacji, sektor zatrudnienia, poziom dochodów, liczba osób pozostających na utrzymaniu użytkownika (i ich wiek), stan posiadania (samochody, nieruchomości etc.), profil etniczny i religijny;

[16] https://en.wikipedia.org/wiki/Do_Not_Track

- szczegółowe dane o lokalizacji ustalone na podstawie współrzędnych GPS, sieci Wi-Fi i adresów IP, z którymi łączyło się urządzenie użytkownika;
- dane techniczne, takie jak rodzaj systemu operacyjnego, ustawienia przeglądarki i rozdzielczość ekranu, które składają się na (coraz bardziej unikatowy) „odcisk palca" urządzenia;
- historia aktywności na stronie: w co użytkownik kliknął, co przykuło jego uwagę (i na jak długo), co ostatecznie kupił (i za ile).

Budując profil możliwego klienta (*look-alike*), agencje mediowe próbują ustalić nawyki, zainteresowania, słabości, ważne momenty z życia (takie jak ślub czy ciąża), cechy osobowościowe i demograficzne użytkowników. Na tej podstawie tworzą wąsko zdefiniowane kategorie, powiązane z cechami usługi czy produktu, który mają za zadanie sprzedać: „aktywny styl życia i SUV", „dom, zdrowe jedzenie", „przede wszystkim dzieci", „miejski styl życia, singiel" czy „ponadprzeciętny dochód, dobra luksusowe". Podobnie działają brokerzy danych, którzy w swoich ofertach dla agencji mediowych i innych nabywców wyodrębniają **segmenty klientów** odpowiadające określonym cechom lub przewidywanym zachowaniom oraz punktację (ang. *score*) wskazującą na prawdopodobieństwo, że takie cechy lub zachowania się potwierdzą.

Facebook w ramach oferty dla swoich klientów (reklamodawców) jest w stanie wygenerować tysiące szczegółowych kategorii związanych z określonymi markami produktów, zainteresowaniami (rodzajem uprawianego sportu, śledzonymi serialami, ulubioną muzyką), konkretnym stylem czy momentem życia (np. para bez dzieci w dużym mieście, świeżo upieczeni rodzice), siłą nabywczą, a nawet precyzyjnie określonymi zamiarami zakupowymi (np. planowany zakup BMW w ciągu 2 miesięcy wraz z budżetem, jaki klient jest skłonny w ten zakup zainwestować).

Dane osobowe zbierane i przetwarzane w celach marketingowych mają różne źródła i charakter. Mogą pochodzić bezpośrednio od użytkowników, jeśli zostały przez nich udostępnione w sposób dobrowolny (np. na etapie zawierania umowy) lub zadeklarowane (np. w ramach badania). Przeciętny użytkownik jest w stanie kontrolować tylko te informacje, choć nawet to wymaga realnego wysiłku (np. uważania na to, jakie treści wrzuca, lajkuje lub wyszukuje w sieci). A to tylko czubek góry lodowej, **pierwsza warstwa cyfrowego profilu**.

Drugą warstwę cyfrowego profilu tworzą wyniki analizy behawioralnej. Bardzo trudno ją kontrolować, ponieważ źródłem tych danych jest ciągłe monitorowanie aktywności użytkowników, często w sposób niejawny lub enigmatycznie wspomniany w ogólnych warunkach świadczenia usług. Na obserwację behawioralną składa się rejestrowanie m.in. wypowiadanych do mikrofonu słów, prędkości pisania, rzeczywistej lokalizacji,

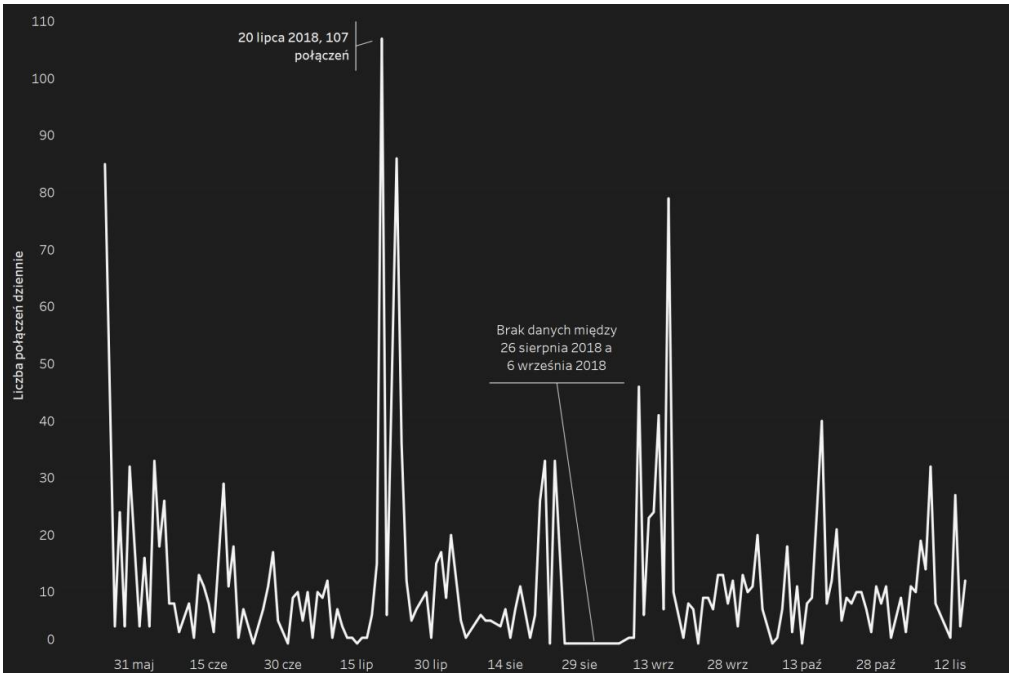
sposobu korzystania z urządzeń mobilnych (od ich pozycji w przestrzeni, przez naładowanie baterii, po nacisk i ruch palca na ekranie), „konsumowanych" treści (od tego, na co klika użytkownik, po czas spędzony w danym miejscu i ruchy kursora na ekranie).

Co wie o Tobie dostawca Twojej poczty elektronicznej?

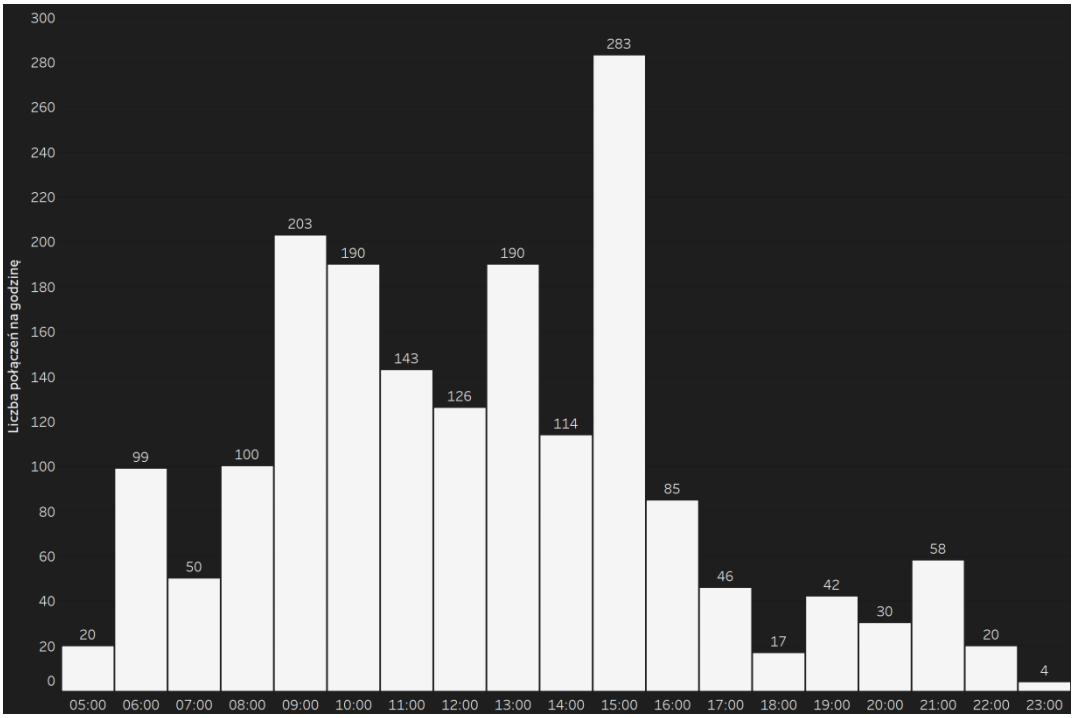
Poniższe wykresy obrazują wybrane dane, które jeden z pracowników Fundacji Panoptykon otrzymał od operatora poczty elektronicznej, z której korzysta.

Connection Time	IP	Type	Browser	Dev model	Manufacturer	Mobile	Hardware	Touch	Diagonal	Screeen	OS	OS ver	Browser	Bver	ScreenWidth	ScreenHeight	Langu
2018-05-25 05:51:55	89.65.44.137	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 7	NT 6.1	Firefox	56	1280	720	null
2018-05-25 05:52:08	89.65.44.137	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 7	NT 6.1	Firefox	56	1280	720	null
2018-05-25 09:21:59	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:22:11	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:22:17	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:22:30	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:22:34	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:22:44	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:31:42	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:32:11	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:35:03	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:35:08	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:35:31	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:35:40	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:35:59	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:04	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:12	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:16	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:21	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:37	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:40	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:42	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:36:53	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:37:11	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:37:13	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:37:36	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:37:41	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 09:37:47	85.222.56.19	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 13:32:16	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	null	null	null
2018-05-25 14:13:13	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:13:26	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:13:31	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:13:34	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:13:47	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:14:07	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:14:12	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null
2018-05-25 14:14:18	89.79.92.104	Mozilla	Firefox - Windows	null	null	0	Desktop	0	null		Windows 10	NT 10.0	Firefox	59	1366	768	null

Tabela przedstawia surowe dane dotyczące aktywności użytkownika: dokładne godziny logowań, numery IP, przeglądarka, z której korzystał, system operacyjny, ustawienia sprzętu. To tylko wybrane informacje, ale nawet pobieżna analiza pozwala wysnuć interesujące wnioski.



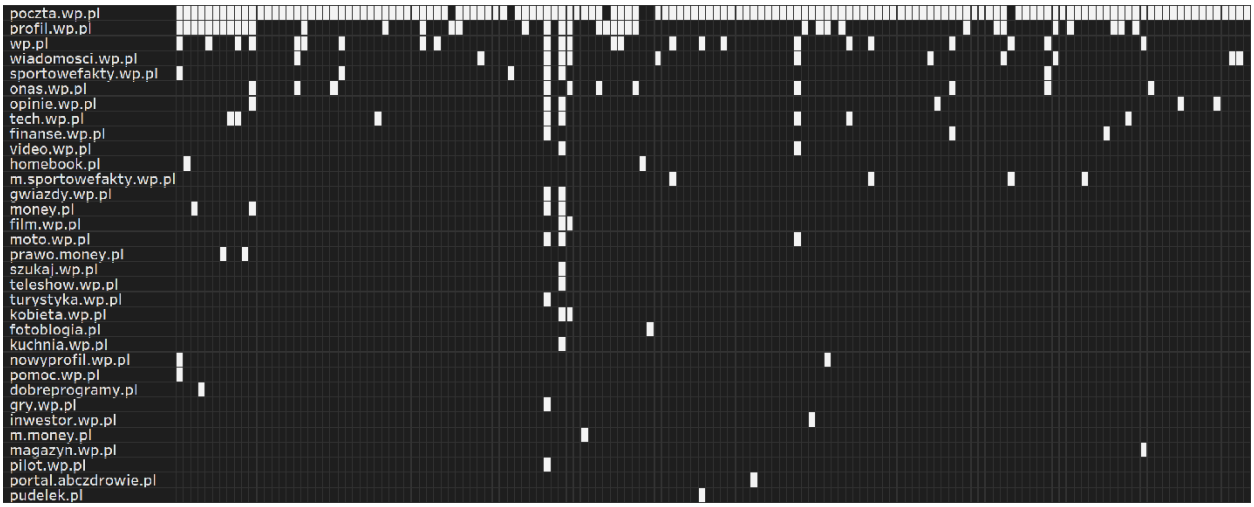
Liczba i godziny logowań w dłuższym okresie tworzą wyraźny wzorzec aktywności użytkowmika: można wysnuć wnioski co do jego nawyków i zaobserwować anomalie. Nasz użytkownik często korzysta z poczty w godzinach pracy (szczyt przypada na 15:00), a brak logowań między 26 sierpnia a 6 września sugeruje, że był wtedy na wakacjach.



Wiemy też, z usług jakich operatorów sieci korzystał:

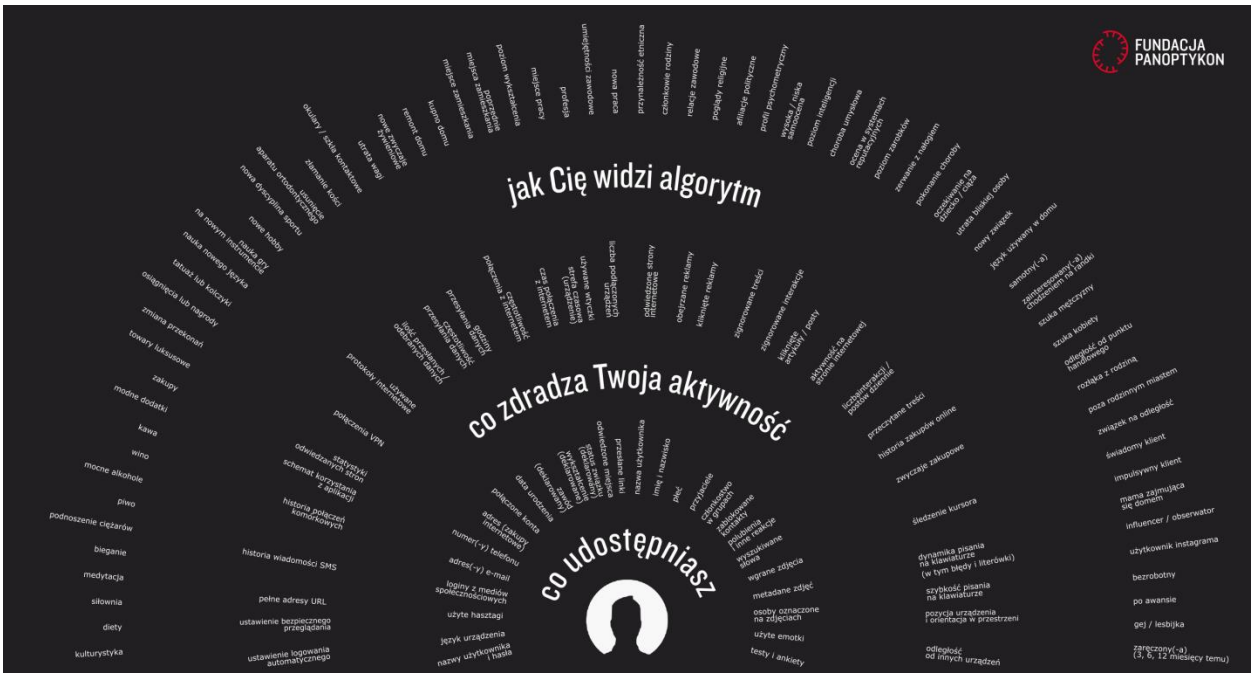
Operator Internetu	
UPC Polska	1 427 połączeń
Plus	194 połączeń
Polkomtel Sp. z o.o.	144 połączeń
Orange Polska	50 połączeń
Play	4 połączeń
T-mobile Polska	1 połączeń

A także, z jakich domen przechodził do poczty najczęściej (wygląda na to, że tylko przez 2–3 dni w trakcie badanego okresu nudził się w pracy lub w domu na tyle, by częściej wchodzić na pocztę z innych stron serwisu):



Analiza danych obejmujących aktywność użytkownika buduje **drugą warstwę cyfrowego profilu**. Nasza próbka danych pochodzi tylko z jednej i to stosunkowo nieinwazyjnej usługi, jaką jest poczta elektroniczna. Wyszukiwarka lub typowa aplikacja mobilna dysponują bez porównania bogatszym zasobem danych behawioralnych.

Trudno mówić o jakiejkolwiek kontroli użytkownika nad **trzecią warstwą cyfrowego profilu**, którą tworzą dane „wywnioskowane” (czy też: „wymodelowane”) w oparciu o analizę statystyczną. To jedynie przypuszczenia na temat ukrytych cech lub przewidywanych zachowań użytkownika, wygenerowane na podstawie danych rzeczywistych. W grę może wchodzić ustalenie poglądów religijnych i politycznych, nawyków zakupowych, wzorców poruszania się, siły nabywczej, relacji rodzinnych, ważnych zmian w życiu, stanu zdrowia, nałogów i najdziwniejszych nawet zainteresowań. Te dane mówią mniej o konkretnej osobie, więcej o „statystycznym użytkowniku” reprezentującym określony styl życia. To jednak nie przeszkadza firmom z branży reklamowej traktować ich poważnie. W praktyce mogą być one najcenniejszą składową cyfrowego profilu.



Otwórz grafikę w nowym oknie, aby ją powiększyć.

4. Ciemna strona śledzenia i profilowania użytkowników sieci

Z perspektywy użytkowników dominujący model generowania zysku w oparciu o eksploatację danych osobowych ma więcej wad niż zalet. Cena, jaką ostatecznie płacą za dostarczane im usługi i treści (notabene coraz gorszej jakości), jest nieproporcjonalnie wysoka i nie zawsze widoczna. Oto nasz przegląd zagrożeń związanych ze śledzeniem i profilowaniem.

Brak przejrzystości i kontroli

Użytkownicy nie mają realnego wpływu na zakres danych, jakie są zbierane lub generowane (na zasadzie predykcji, w oparciu o korelacje statystyczne) na ich temat, szczególnie przez tzw. strony trzecie (np. brokerów danych). Tym bardziej nie mają wpływu na kryteria profilowania, któremu są poddawani, ani na dobór treści, jakie ostatecznie zobaczą na ekranie swojego urządzenia.

Powszechnie wiadomo, że za dobór wyświetlanych treści odpowiadają uczące się algorytmy i sieci neuronowe[17]. Na tym publiczna dyskusja się kończy, pozostawiając zbyt duże pole do domysłów dotyczących logiki działania tych systemów oraz kryteriów profilowania (w tym tego, w jakim stopniu są to „decyzje redakcyjne" samych firm). Nawet wiodące firmy technologiczne, takie jak Alphabet/Google i Facebook, które w ramach swoich platform udostępniły użytkownikom interfejsy do zarządzania profilami reklamowymi, nie ujawniają pełnego zakresu danych przetwarzanych w celach marketingowych (w szczególności pozyskiwanych od tzw. stron trzecich, np. brokerów danych) ani logiki stojącej za profilowaniem wyświetlanych treści[18].

Na ten problem zwrócił uwagę niemiecki minister ds. ochrony konsumentów Heiko Maas[19]: „Brak przejrzystości to nasz wspólny problem. Wiemy, że niezliczone dane osobowe przepływają przez Internet, że te dane mogą być łączone i analizowane. Ale szczegóły tych transakcji pozostają dla nas niedostępne. Kto wie, jakie dane rzeczywiście przesyłają dalej nasze smartfony? Kto liczy się z tym, że nawet rytm, w jakim przesuwamy palcami po klawiaturze, może zdradzić, w jak bardzo konsumpcyjnym nastroju właśnie jesteśmy?

---

[17] Więcej na ten temat: Fundacja Panoptykon, *Prawda algorytmów*, https://panoptykon.org/wiadomosc/prawda-algorytmow.

[18] Fundacja Panoptykon, *Targetowana reklama to nadal czarna skrzynka*, https://panoptykon.org/wiadomosc/targetowana-reklama-nadal-czarna-skrzynka.

[19] Heiko Maas, przemówienie na konferencji *Digital live – networked. measured. Sold? #values #Algorithms #IoT*, 3 lipca 2017, Berlin.

Kto jest w stanie przewidzieć, że zdjęcia, jakie umieszcza na swoim Instagramie, mogą być użyte do oceny jego stanu emocjonalnego?". **W efekcie użytkownicy urządzeń mobilnych i usług internetowych są spychani do roli przedmiotu nieprzejrzystych i nie zawsze uczciwych transakcji.**

Brak przejrzystości i kontroli po stronie użytkowników na wszystkich etapach budowania ich cyfrowych osobowości przekłada się na konkretne ryzyka. W grę wchodzi nie tylko dotkliwe ograniczenie autonomii informacyjnej, będącej jednym z praw podstawowych, ale także ryzyko finansowe lub utrata niematerialnych korzyści. Obecny model zbierania i eksploatacji danych nie gwarantuje, że wygenerowane profile osobowe będą prawdziwe, a jeśli nawet będą – czy nie zostaną wykorzystane w sposób dyskryminujący lub wykluczający użytkownika[20].

Dyskryminacja cenowa i wykluczenie

Wielokrotnie przeprowadzane eksperymenty naukowe i dziennikarskie potwierdzają, że w sferze usług internetowych dochodzi do **dyskryminacji cenowej**[21]. A więc te same produkty (np. laptop, bilet lotniczy) lub usługi (np. abonament telefoniczny) są oferowane konsumentom na tym samym rynku po różnych cenach, uzależnionych od ich unikatowych cech (np. klasy urządzenia, z którego korzystają; ofert, jakie przeglądali wcześniej w sieci, czy ustalonego profilu osobowościowego).

Firmy standardowo wykorzystują systemy CRM (Customer Relationship Management) po to, by skoncentrować uwagę swoich pracowników na utrzymaniu najcenniejszych klientów. W skrajnych scenariuszach takie działania mogą prowadzić do **wykluczenia klientów uznanych za mniej wartościowych** poprzez ograniczenie kierowanych do nich ofert czy intencjonalne utrudnienia w procesie obsługi (np. długie czasy oczekiwania na infolinii, brak dostępnych terminów w placówkach banków). Znane są również przykłady kampanii marketingowych obliczonych na **wykorzystanie słabości lub trudnej sytuacji życiowej** klientów dysponujących realnym kapitałem, np. mieszkaniem na sprzedaż[22].

---

[20] Jędrzej Niklas, *Dyskryminacja w świecie rządzonym przez dane*, https://panoptykon.org/wiadomosc/dyskryminacja-w-swiecie-rzadzonym-przez-dane.

[21] Por. raport Executive Office of the President of the United States, *Big Data and Differential Pricing*, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf;

D. Keats Citron, F. A. Pasquale, *The Scored Society: Due Process for Automated Predictions*; *Data And Discrimination: Collected Essays*, https://na-production.s3.amazonaws.com/documents/data-and-discrimination.pdf, The Wall Street Journal, *Websites Vary Prices, Deals Based on Users' Information*, https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

[22] Por. D. Dudek, *Jak znaleźć zadłużonego właściciela mieszkania, który chce je sprzedać? Możliwości ultraprecyzyjnego targetowania reklamy na Facebook*, http://jakrobicmarketing.pl/jak-znalezc-zadluzonego-wlasciciela-mieszkania-ktory-chce-je-sprzedac-mozliwosci-ultraprecyzyjnego-targetowania-reklamy-na-facebook/.

Zręczne wykorzystanie takiej informacji w połączeniu z ukrytymi technikami marketingowymi może doprowadzić do zawarcia nieuczciwej transakcji.

Krzywdzący scoring[23]

Uzasadnione kontrowersje wzbudzają także metody oceny ryzyka, w coraz większej mierze bazujące na modelach predykcyjnych i analizie korelacji statystycznych (wykorzystujących np. informacje o miejscu zamieszkania, pochodzeniu czy sieci osobistych relacji), a nie na faktycznej historii zachowań i zweryfikowanych cechach ocenianych osób. Surowe dane są wtłaczane w algorytmy, których zadaniem jest kalkulacja ryzyka kredytowego pod kątem przyszłych finansowych lub życiowych ruchów klienta.

W 2013 r. Amerykańska Federalna Komisja Handlu podała, że aż 10 mln osób miało w swoich raportach historii kredytowej błędy na tyle poważne, by w efekcie podwyższyć koszty kredytu. W Niemczech działanie modeli scoringowych badał projekt OpenSCHUFA, do którego zgłosiło się 20 000 wolontariuszy – ludzi, którzy zdecydowali się zweryfikować swój scoring. Porównując wyniki ocen scoringowych z danymi finansowymi, które wolontariusze dobrowolnie udostępnili, badacze stwierdzili, że statystyczne modele stosowane przez banki prowadzą do błędnych i niesprawiedliwych decyzji.

Tymczasem na rynku amerykańskim standardem staje się wykorzystywanie w scoringu niestandardowych kategorii danych, np. danych behawioralnych (sposobu wypełniania formularza; prędkości podejmowania decyzji; liczby popełnionych i skorygowanych błędów; urządzenia, z jakiego korzystał klient, i tego, czy miało ono naładowaną baterię). **W efekcie instytucje finansowe i ubezpieczeniowe zaczynają podejmować istotne dla swoich klientów decyzje w oparciu o niezweryfikowane i dyskusyjne podstawy**. Jednocześnie rośnie wpływ złego scoringu: życie takiej informacji nie kończy się na pojedynczej odmowie kredytu czy ubezpieczenia, ale może obciążyć profil danej osoby w relacji z organami państwa, pracodawcą czy partnerami biznesowymi.

Manipulowanie emocjami[24]

Kolejny obszar, w którym użytkownicy usług internetowych doświadczają utraty kontroli i obawiają się manipulacji, to dostęp do informacji w postaci wyników wyszukiwania czy treści wyświetlanych przez portale społecznościowe (Facebook, Twitter, Instagram).

---

[23] Więcej na ten temat: Fundacja Panoptykon, *Dlaczego bank odmówił Ci kredytu?*, https://panoptykon.org/wiadomosc/dlaczego-bank-odmowil-ci-kredytu.

[24] Więcej na ten temat: Fundacja Panoptykon, *Prześwietleni przez algorytm*, https://panoptykon.org/wiadomosc/przeswietleni-przez-algorytm.

W 2014 r. wyszło na jaw, że 700 tys. użytkowników Facebooka wzięło udział – nieświadomie – w eksperymencie badającym wpływ negatywnych i pozytywnych komunikatów na ich zachowanie w sieci. To badanie, przeprowadzone bez wiedzy i zgody użytkowników, wywołało falę niezadowolenia i protesty. A to był tylko początek. W 2017 r. za sprawą wycieku informacji handlowych z biura Facebooka w Australii opinia publiczna dowiedziała się, że portal pozwala na profilowanie reklam i innych targetowanych przekazów w oparciu o stan emocjonalny użytkowników (nawet tych 14-letnich!). Według doniesień dziennika The Australian oferta skierowana do partnerów biznesowych obejmuje szerokie spektrum stanów emocjonalnych nastolatków: od „niepewny swojej wartości", „zagrożony", „beznadziejny" i „głupi" po „nieudaczny", „zestresowany" czy „przechodzący życiowy kryzys".

Wyborcza dezinformacja[25]

Więcej światła na ryzyka związane z targetowaniem przekazu, nie tylko reklamowego, w mediach społecznościowych rzuciła dyskusja, jaka przetoczyła się po wyborze Donalda Trumpa na prezydenta USA i po brytyjskiej kampanii Leave.EU. Mimo że oba sztaby korzystały z od dawna znanych i sprawdzonych technik marketingowych, po raz pierwszy na taką skalę zainteresowała się nimi opinia publiczna. Śledztwa dziennikarskie i doniesienia byłego dyrektora ds. badań Cambridge Analytica Christophera Wyliego dowodziły, że mieliśmy do czynienia z **szeroko zakrojoną i precyzyjnie wyreżyserowaną manipulacją**, w dużej mierze za sprawą mikrotargetowania użytkowników Facebooka i taktycznego wykorzystania *fake news*.

Te zarzuty w dużej mierze potwierdzali architekci obu kampanii, publicznie chwaląc się możliwościami **niejawnego wpływania na poglądy i postawy wyborców** m.in. poprzez *dark posts* (treści widoczne tylko dla konkretnych użytkowników, intensywnie wykorzystywane do ich zmobilizowania lub zdemobilizowania w dzień głosowania). Na ile te działania przesądziły o wyniku wyborów, pozostaje kwestią spekulacji. Faktem jest, że Facebook dużo obiecuje specjalistom od marketingu politycznego, oferując im, nawet w Europie, przeznaczone dla nich usługi.

—

[25] Więcej na ten temat: Fundacja Panoptykon,*Polityczny marketing – nieoczekiwane zagrożenie dla demokracji?*, https://panoptykon.org/wiadomosc/polityczny-marketing-nieoczekiwane-zagrozenie-dla-demokracji.

O Fundacji Panoptykon

Kto zna Cię najlepiej? Rodzina? Przyjaciele? Po prostu Ty?

A co, jeżeli więcej wie o Tobie megakorporacja z Doliny Krzemowej lub państwo, które zapukało do niej po Twój cyfrowy profil? I co, jeżeli ta wiedza służy temu, by zarobić na Twoich słabościach, wpłynąć na Twoje wybory lub przewidzieć, co zrobisz?

Nie możesz odciąć się od swojego cyfrowego profilu, nie możesz go wykasować. Ale to nie znaczy, że nie możesz nic zrobić.

Fundacja Panoptykon patrzy na ręce państwu i firmom. Sprawdzamy, jak wykorzystują Twoje dane, ujawniamy nadużycia, walczymy o prawo chroniące wolność i prywatność. Pokazujemy, jak świadomie poruszać się w coraz bardziej cyfrowym świecie.

Dowiedz się więcej na panoptykon.org i pomóż Fundacji Panoptykon kontrolować kontrolujących!

panoptykon.org/wspieraj

Przekaż 1% podatku. KRS: 0000327613

Słowniczek

AAID	unikatowy numer identyfikacyjny użytkownika urządzenia mobilnego z systemem Android, wykorzystywany do celów reklamowych (z ang. *Android Advertising ID*)
Ad tech	zbiorcze określenie różnego rodzaju narzędzi i oprogramowania wykorzystywanego w branży reklamy internetowej
Bid request	komunikat wysyłany przez platformę podaży i giełdę reklam, sygnalizujący platformom popytu (reklamodawcom), że impresja jest dostępna do licytacji; wraz z komunikatem przesyłane są takie dane jak: identyfikator użytkownika, link strony, na której załadowanie użytkownik czeka, lokalizacja, a czasem także wiek, płeć i zainteresowania użytkownika
Broker danych	pośrednik w handlu danymi, który zdobywa dane o użytkownikach Internetu za pośrednictwem innych firm (banków, sklepów, ubezpieczycieli, innych stron internetowych etc.); wspólną cechą brokerów danych jest to, że użytkownicy nie stykają się z nimi bezpośrednio
Browser fingerprint	tzw. odcisk palca przeglądarki, metoda pozwalająca na jednoznaczne oznaczenie przeglądarki internetowej i śledzenie użytkownika nawet wtedy, gdy wyłączył ciasteczka
***Cookie* (ciasteczko)**	niewielki plik tekstowy zapisywany w pamięci komputera przez przeglądarkę, otwierany przy następnych wejściach na stronę; pliki *cookie* mogą mieć różne funkcje: najczęściej są wymagane do sprawnego działania serwisu, czasem ich podstawową funkcją jest śledzenie użytkowników na różnych portalach; jednym z elementów zapisanych w pliku *cookie* jest nazwa konkretnej domeny, co powoduje, że może on zostać odczytany tylko przez ten podmiot, który go zapisał

Cookie syncing	proces synchronizowania (wymiany) plików *cookie* pomiędzy poszczególnymi pośrednikami po zakończeniu licytacji na giełdzie reklam, funkcjonujący również pod nazwą *cookie matching*
Dark patterns	w wolnym tłumaczeniu: „wredne praktyki", elementy interfejsów w aplikacjach i serwisach internetowych zaprojektowane tak, by skłonić użytkownika do wybrania opcji najkorzystniejszej dla firmy zarabiającej na komercjalizacji danych
***Data Management Platform* (DMP)**	platforma zarządzania danymi, tj. wyspecjalizowane oprogramowanie, którego funkcją jest katalogowanie identyfikatorów i profili od różnych pośredników
***Demand Side Platform* (DSP)**	platforma popytu, tj. oprogramowanie pomagające reklamodawcom kupić reklamę na giełdzie reklam i wyświetlić ją odpowiednio sprofilowanemu użytkownikowi; w tym celu odbiera i analizuje dane wystawiane przez sprzedawców przestrzeni reklamowej i składa oferty podczas licytacji w modelu *real-time bidding*
Exif	metadane zapisanych na urządzeniu zdjęć, które standardowo zawierają m.in. współrzędne geograficzne ustalone w momencie robienia zdjęcia, rozmiar zdjęcia, rodzaj oświetlenia i rodzaj aparatu
Giełda reklam	z ang. *ad exchange*, wyspecjalizowane oprogramowanie pośredniczące w transakcjach reklamowych pomiędzy stroną podaży (wydawca strony internetowej) i stroną popytu (reklamodawca) w modelu *real-time bidding*
IDFA	z ang. *Identifier For Advertising*, unikatowy numer identyfikacyjny użytkowników urządzeń mobilnych Apple wykorzystywany do celów reklamowych
Impresja	z ang. *impression*, pojedyncze wyświetlenia reklam przez potencjalnych odbiorców – użytkowników o określonym profilu, którzy w danym momencie przeglądają stronę

Internet rzeczy	z ang. *Internet of Things*, IoT; dynamicznie rozwijający się sektor gospodarki, w którym fizyczne przedmioty gromadzą, przetwarzają lub wymieniają dane za pośrednictwem Internetu; do tego typu przedmiotów zaliczają się m.in. urządzenia gospodarstwa domowego (smart dom), akcesoria (np. *fitness tracker*) czy zabawki
Look-alike	w wolnym tłumaczeniu: „podobni odbiorcy"; użytkownicy posiadający podobne cechy lub zainteresowania jak użytkownicy należący do już istniejącej grupy odbiorców (np. klientów sklepu); *look-alike* to mieszanka twardych danych pochodzących od reklamodawców i statystycznej wiedzy o ludziach: komponent statystyczny pochodzi z analizy danych zebranych przez dostawców treści, mediów społecznościowych i brokerów danych i wystawionych na sprzedaż agencjom mediowym
Mikrotargetowanie	precyzyjne dobranie indywidualnego przekazu do określonej osoby; w porównaniu do zwykłego targetowania mikrotargetowanie sięga głębiej, wykorzystując nierzadko cechy osobowości; polega na skierowaniu do danej osoby takiego komunikatu, który najlepiej do niej przemówi
Profilowanie	zautomatyzowany proces analizy zachowań użytkownika prowadzący do ocenienia jego osoby lub do wnioskowania o posiadaniu przez niego określonych cech
Programmatic	technologia służąca do automatyzowania sprzedaży i zakupu reklamy w Internecie
Real-time bidding	w wolnym tłumaczeniu: „licytacje w czasie rzeczywistym", odbywające się w sposób zautomatyzowany na giełdzie reklam przy uczestnictwie różnego rodzaju pośredników; przedmiotem licytacji są tzw. impresje, czyli pojedyncze wyświetlenia konkretnych reklam; jedna licytacja w tym modelu trwa ok. 200 milisekund

Reklama behawioralna profilowana targetowana	trzy określenia na ten sam typ reklamy: dopasowanej nie do treści strony, ale do preferencji użytkownika, który ją odwiedza
Reklama internetowa	każdy format reklamowy wykorzystywany w Internecie; najpopularniejsze są reklamy typu *display* (np. graficzne banery), reklamy w wyszukiwarkach oraz e-mail marketing
Reklama kontekstowa	reklama, która jest dopasowana do treści strony, na której się wyświetla (np. w artykule o samochodach użytkownik zobaczy reklamy konkretnych marek)
Segment	sprecyzowana, jednorodna kategoria konsumentów odpowiadająca określonym cechom lub przewidywanym zachowaniom; segmentacja ułatwia dopasowanie klienta do konkretnej kampanii reklamowej
Skrypt śledzący	program, który uruchamia się na stronie internetowej i inicjuje zbieranie danych o użytkownikach
***Supply Side Platform* (SSP)**	platforma podaży, tj. oprogramowanie, za pomocą którego dostawcy treści internetowych wystawiają na sprzedaż profile swoich użytkowników oraz wyświetlaną im przestrzeń reklamową
Targetowanie	kierowanie komunikatu dopasowanego do konkretnego odbiorcy posiadającego cechy, które mogą sprawić, że będzie zainteresowany danym produktem
Zaufani partnerzy	firmy z ekosystemu reklamy internetowej, z którymi współpracuje wydawca i w których imieniu zbiera zgodę użytkownika na śledzenie i profilowanie

Polecane źródła:

Cracked Labs, *Report – Corporate Surveillance in Everyday Life*, http://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf.

Interactive Advertising Bureau, *Open RTB 3.0 Beta specification*, https://www.iab.com/guidelines/real-time-bidding-rtb-project/.

Me and My Shadow, *Location tracking*, https://myshadow.org/location-tracking.

The Norwegian Consumer Council, *Appfail Report – Threats to Consumers in Mobile Apps*, https://www.forbrukerradet.no/undersokelse/2015/appfail-threats-to-consumers-in-mobile-apps/.

The Norwegian Consumer Council, *Deceived by Design. How tech companies use dark patterns to discourage us from exercising our rights to privacy*, https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf.

Share Labs, *Invisible Infrastructures: Mobile permissions*, https://labs.rs/en/invisible-infrastructures-mobile-permissions/.