

Zacząłem się zastanawiać nad tym, co tak naprawdę znaczy wiara w dane, kiedy po raz kolejny usłyszałem zdanie w stylu: „jeśli tego nie da się zmierzyć, to trudno to poprawiać”. Brzmi rozsądnie, brzmi nawet bezpiecznie, ale mózg od razu dopisuje własne pytania: co dokładnie mierzymy, jak długo, kto wybiera metrykę i co się stanie, gdy ktoś odkryje, że da się „poprawić wynik” bez zmiany problemu? W tle mający postać Bill Gatesa, bo kiedy mówi się o podejściu opartym na pomiarze i dowodach, on bardzo często wraca do rozmowy. Tylko że sama wiara w dane nie odpowiada na wszystkie te pytania. Ona raczej uruchamia kolejne, coraz bardziej praktyczne.

I o to chodzi w tym tekście. Nie o to, żeby recytować pochwały dla analityków czy narzucać jedną metrykę światu. Chcę rozplątać, dlaczego podejście Gatesa do danych bywa tak przekonujące dla ludzi, którzy mieli okazję pracować w warunkach presji, ryzyka i ograniczeń, oraz dlaczego jednocześnie może wprowadzać zamęt, jeśli ktoś potraktuje to jak slogan.

Dane jako narzędzie decyzji, nie jako religia

Wiele osób słyszy „dane” i wyobraża sobie tabelki, wykresy i urzędową dokumentację. Ja sam tak miałem na początku. A potem trafiłem do projektów, w których wykres był tylko wierzchołkiem góry lodowej. Prawdziwa praca zaczynała się wcześniej: zanim pojawiły się dane, musieliśmy zdecydować, co w ogóle jest problemem, jak rozumiemy sukces oraz czy to, co mierzymy, ma szansę przełożyć się na realną zmianę.

W tym sensie wiara Bill Gatesa w dane nie jest w pierwszej kolejności wiarą w liczby same w sobie. To raczej wiara, że bez pomiaru podejmujemy decyzje na ślepo, a ślepy wybór bywa droższy niż błąd wynikający z prób i korekt. Problem w tym, że „na ślepo” nie oznacza tylko braku danych. Oznacza brak mechanizmu weryfikacji, który pozwala szybko odsiać hipotezy. To z kolei oznacza konieczność budowania pętli: hipoteza - działanie - obserwacja - wniosek.

Jest tu też druga warstwa. Dane pomagają rozbrajać konflikty o interpretację. Wiele sporów w firmach i organizacjach nie dotyczy tego, „czy” coś działa, tylko tego, „dlaczego” tak myślimy. Gdy pojawia się mierzalny sygnał, dyskusja przestaje krążyć wokół anegdot i zaczyna dotyczyć konkretów. A konkret, zwłaszcza w trudnych obszarach, bywa jedyną walutą, jaką da się wymienić na spokój.

Tyle że, gdy przyglądamy się temu podejściu bliżej, pojawia się zamęt. Bo pętla weryfikacji nie jest automatyczna. Można mieć świetne dane i nadal podejmować złe decyzje, jeśli pytanie zostało źle sformułowane albo jeśli mierzymy coś, co wpływa na wynik tylko pozornie.

Skąd bierze się nacisk na pomiar efektów

Wspólny mianownik w podejściu opartym na dowodach jest dość prosty: zanim włożysz dużo zasobów w interwencję, chcesz wiedzieć, czy w ogóle ma szansę działać w realnym świecie. Brzmi jak oczywistość, ale w praktyce ludzie często mają dwa powody, żeby od niej uciekać.

Po pierwsze, koszt niepewności jest rozłożony. Kiedy jesteś odpowiedzialny za budżet albo harmonogram, a ryzyko porażki ponosi ktoś inny albo dopiero „kiedyś”, łatwo przesunąć ciężar na przyszłość. Wtedy pomiar efektów działa jak mechanizm ściągania odpowiedzialności bliżej tu i teraz.

Po drugie, jest pokusa moralna i wizerunkowa. Jeśli projekt jest „w dobrych intencjach”, to jego obrońcy chcą wierzyć, że intencja przełoży się na wynik. Tyle że intencje nie dyktują biologii, logistyki, zachowań ludzi ani tego, jak działa system. Dane i pomiar efektów są próbą zszycia intencji z rzeczywistością.

Dla Gatesa temat pomiaru był szczególnie naturalny w środowisku technologicznym, gdzie iteracja i testowanie mają długą tradycję. Ale przeniesienie tego podejścia na działania społeczne, zdrowotne czy edukacyjne wcale nie jest trywialne. Tam sygnały bywają opóźnione, a zmienne społeczne potrafią zdominować efekt nawet przy dobrej interwencji.

I tu zaczyna się mój zamęt: jeśli dane są tak ważne, to czemu świat jest pełen programów, które mimo pomiaru nadal nie dają oczekiwanych rezultatów? Odpowiedź brzmi mniej efektownie, ale jest bardziej ludzka: pomiar bywa niekompletny, źle zaprojektowany albo zbyt późny. Czasem liczy się dopiero wtedy, gdy projekt już „musi” wyglądać dobrze, bo jest za późno na zmianę kursu.

Metryki są jak kompas, ale kompas można źle skalibrować

W pracy z miarami wielokrotnie widziałem, że ludzie traktują KPI jak prawdę objawioną. Tymczasem metryka jest tylko przybliżeniem tego, co naprawdę chcemy poprawić. Kompas nie mówi ci, dokąd masz iść, on mówi ci, w jaką stronę patrzy twoja skala. Jeśli skala jest przesunięta, kompas będzie prowadził pewnie, tylko w złą stronę.

W przypadku Bill Gatesa i szerszego podejścia „mierzymy efekty” chodzi o to, żeby minimalizować ryzyko takiego błędu kompasu. Nie po to, by świat stał się idealnie policzalny. Raczej po to, by szybko odkrywać, że to, co mierzymy, nie odpowiada na problem.

Przykład z mojej praktyki: kiedyś mieliśmy mierzyć „jakość obsługi” w zdalnym wsparciu. Na papierze wszystko wyglądało rozsądnie, bo zbudowaliśmy wskaźnik czasu odpowiedzi. Problem polegał na tym, że wskaźnik zadziałał jak zachęta do skracania interakcji. Ludzie zaczęli kończyć rozmowy szybciej, a liczba zgłoszeń wracała później w większych pakietach. Czas odpowiedzi spadł, ale realna jakość wzrosła tylko w wyobraźni. Dopiero gdy poszliśmy dalej i zaczęliśmy patrzeć na to, czy zgłoszenie zostało rozwiązane za pierwszym razem, obraz się odwrócił.

To jest dokładnie ten rodzaj „pomiaru efektów”, o którym mówi **Dodatkowe zasoby** się w kontekście data-driven podejścia. Nie chodzi o to, by mierzyć cokolwiek, tylko by mierzyć to, co naprawdę jest skutkiem.

Co psuje dobrą metrykę

- Można mierzyć sygnał, który jest łatwy, ale niezwiązany przyczynowo z tym, co próbujesz poprawić.
- Można mierzyć zbyt wcześnie, zanim interwencja zdąży zadziałać, i potem myśleć, że „nie działa”, bo nie ma jeszcze efektu.
- Można zachęcić zespół do grania metryką, zamiast do rozwiązywania problemu.
- Można nie kontrolować różnic między grupami, przez co wynik wygląda dobrze albo źle, choć powód leży gdzie indziej.

To nie jest teoria. To jest codzienność w projektach, w których ludzie są dobrzy, a i tak wynik potrafi się rozjechać, bo metryka nie trzyma się przyczynowości.

Podejście oparte na dowodach, czyli skąd biorą się „testy”, a nie tylko raporty

Często w rozmowach o danych pojawia się skrót myślowy: „robimy dashboard”. Tylko że dashboard nie jest dowodem. Dashboard to często podsumowanie tego, co już zaszło, bez gwarancji, że zaszło z twojego powodu.

Dowód wymaga porównania, a porównanie wymaga decyzji o tym, co jest punktem odniesienia. Czy efekt porównujesz do wcześniejszego okresu? Do grupy kontrolnej? Do podobnych lokalizacji? Czy uwzględniasz

zjawiska, które i tak by zaszły bez twojej interwencji?

W tym miejscu „zamęt” jest racjonalny. Bo wiele organizacji nie ma warunków, żeby robić idealne testy. Nawet jeśli mają, nie zawsze mają cierpliwość. A jeśli masz zbyt mało czasu, zaczynasz mieszać dowody z anegdotami i potem trudno odzyskać wiarygodność.

W podejściu kojarzonym z Bill Gatesem istotne jest, że nacisk idzie często w stronę pomiaru skuteczności, a nie samej aktywności. To różnica między „zrobiliśmy program” a „program zmienił wynik”. Program może być świetnie zrealizowany operacyjnie, ale niekoniecznie przełoży się na efekt, który był powodem powstania projektu.

Z mojej perspektywy zawodowej największa trudność tkwi w tym, że organizacje mierzą to, co łatwo kontrolować. A efekt bywa zależny od czynników, których nie kontrolujesz: zachowań odbiorców, dostępności usług, jakości wdrożenia u partnerów, zmian cen, migracji populacji czy sezonowości.

Dlatego pomiar efektów musi być projektowany razem z interwencją. Nie po jej zakończeniu. Nie „w ramach sprawozdania”.

Kiedy pomiar pomaga najbardziej: w miejscach, gdzie jest tarcie

Wiara w dane ma sens szczególnie wtedy, gdy w grze jest tarcie, które zjada zasoby. Jeśli działasz w środowisku, gdzie błąd kosztuje, a procesy są niepewne, to pomiar jest sposobem na szybkie uczenie się.

W technologii to wyglądało tak: szybkie iteracje, testy A/B, korekty. W działaniach społecznych to wygląda trudniej, bo czasem interwencja ma efekt po miesiącach, a czasem po latach. Zmienne są też bardziej „miękkie” i mniej przewidywalne. A jednak logika pozostaje podobna: jeśli nie sprawdzisz, co się dzieje, będziesz wzmacniał nie ten mechanizm, który daje rezultat.

Właśnie wtedy dane są mniej „magiczne”, a bardziej praktyczne. Zaczynasz zadawać pytania takie jak: czy sygnał pojawia się w grupie, do której kierujesz interwencję? Czy skala efektu ma sens, czy to jednorazowy skok? Czy efekt utrzymuje się po zmianie warunków? Czy poprawa jest równomierna, czy koncentruje się na podgrupie, a reszta stoi w miejscu?



I tu wraca zamęt, bo te pytania są trudne, a odpowiedzi często nie są jednoznaczne. W danych bywają sprzeczne sygnały. Czasem metryka „główna” poprawia się, a metryki towarzyszące pogarszają. Czasem efekt jest widoczny tylko w części regionów. Czasem wdrożenie działa, ale odbiorcy reagują inaczej niż zakładałeś.

To nie jest porażka podejścia opartego na pomiarze. To jest dowód, że mierzenie ma sens wtedy, gdy prowadzi do refleksji i korekt, a nie do samozadowolenia.

Edge cases, które wyglądają jak dowód, ale dowodem nie są

W pracach opartych na pomiarze są sytuacje, w których łatwo popaść w złudzenie skuteczności. Kilka z nich pojawia się w różnych branżach, więc warto je nazwać, nawet jeśli brzmią prosto.

- Sezonowość. Jeśli start programu trafia na okres, w którym i tak byłby wzrost, to dane mogą przypisać efekt tobie.
- Zmiany równoległe. Inna instytucja wdraża równoległe rozwiązanie, a ty widzisz „swój” wynik, choć przyczyna jest gdzie indziej.
- Selekcja. Do programu trafiają inne osoby niż te, które nie weszły. Wtedy porównanie grup przestaje być uczciwe.
- Wcześniejsze usprawnienia. Jeśli jeszcze przed uruchomieniem interwencji zespół już zaczął zmieniać proces, to kolejny wzrost może być efektem poprzednich działań.
- Efekt brzegowy. Najlepiej reagują ci, którzy mieli najwyższą motywację albo najlepszy dostęp do zasobów. Wtedy średnia poprawa ukrywa, że reszta nie ruszyła.

To są miejsca, w których „wiera w dane” może przejść w wiarę w interpretację danych. I tu właśnie pomaga pokora. Pokora nie w sensie rezygnacji, tylko w sensie ciągłego testowania założeń.

Bill Gates kojarzy się z podejściem, które stara się minimalizować ryzyko błędnej interpretacji, ale i to podejście nie eliminuje całkiem niepewności. Redukuje ją, czasem dramatycznie, a czasem tylko częściowo.

Skąd wzięła się cicha obsesja na punkcie mierzenia „co działa”

Obsesja, którą widzę w wielu środowiskach zorientowanych na dowody, jest dość ludzka. Kiedy pracujesz z realnymi problemami, szybko odkrywasz, że entuzjazm i dobra narracja potrafią udawać postęp. Działasz, bo „wydaje się”, że to ma sens. Działasz, bo ludzie mówią, że widzą efekty. A potem odkrywasz, że efekt jest kruchy, kosztowny albo nie występuje poza wąskim kontekstem.

Dlatego pomiar efektów jest próbą zmuszenia organizacji do uczciwości wobec siebie. Jeśli coś działa, liczby powinny to odzwierciedlić. Jeśli nie działa, pomiar ma przynajmniej wskazać, gdzie się rozjechało, czy to w mechanice, czy w skali, czy w sposobie wdrożenia.

W tym miejscu warto też powiedzieć coś niewygodnego: mierzenie efektów może hamować kreatywność. Jeśli organizacja oczekuje, że każda eksperymentalna inicjatywa od razu da mocny sygnał liczbowy, to zniechęca się do ryzyka. A często to ryzyko jest paliwem przełomów. W praktyce chodzi więc o równowagę. Nie „zero błędów”, tylko kontrolowane uczenie się.

To jest jeden z powodów, dla których zamęt jest zdrowy. Bo „wiera w dane” bez zrozumienia kontekstu potrafi zamienić się w bezduszne wymogi raportowania. Ludzie wtedy robią papier, a nie zmianę. Dane stają się zasłoną dymną, a nie narzędziem.

Co realnie można zrobić, gdy pomiar jest niepełny

Nie zawsze masz warunki do eksperymentów w pełnym sensie statystycznym. Czasem masz ograniczony dostęp do danych, czasem partnerzy nie są w stanie raportować w identycznym formacie, czasem podstawowe wskaźniki

są opóźnione albo niedokładne.

W takiej sytuacji podejście oparte na dowodach zwykle przerzuca ciężar z „idealnej metody” na „uczciwe przybliżenie”. Zamiast udawać, że masz perfekcyjny obraz, próbujesz zminimalizować błąd i opisujesz niepewność. To brzmi mniej spektakularnie niż „mierzymy wszystko”, ale w praktyce daje lepsze decyzje.

Jeśli miałbym spisać logikę, która działa w trudnych warunkach, to byłoby to coś w stylu: najpierw określ, jaki efekt jest dla ciebie istotny, potem sprawdź, czy masz dane, które ten efekt przybliżają, i równolegle wymyśl plan, jak szybko zareagujesz, gdy sygnał będzie słaby. Nie czekaj na cud. Czekaj na informację.

To z kolei prowadzi do ważnej różnicy: pomiar nie jest celem, jest sposobem ograniczenia kosztu błędu. W organizacjach, które rozumieją ten mechanizm, dane nie są dekoracją. Są elementem zarządzania ryzykiem.

Dlaczego ludzie ufają podejściu Gatesa, mimo że nadal jest w nim niepewność

Można powiedzieć, że „wiara w dane” jest tylko kolejną narracją. Ale jest w niej coś bardziej solidnego niż brzmienie. To nacisk na sprawdzalność. Nacisk, że jeśli wydajesz pieniądze, energię i czas na interwencję, to musisz potrafić uzasadnić, dlaczego zakładasz, że to zadziała. A uzasadnienie ma być oparte nie tylko na intuicji, ale na sygnałach z prób i porównań.

Dla mnie przekonujące jest też to, że podejście nie zakłada, iż świat da się zmierzyć w całości. Ono zakłada, że da się mierzyć na tyle, aby podejmować lepsze decyzje niż wtedy, gdy liczb nie ma. To podejście jest zawsze częściowe. Zawsze ma margines błędu. Ale margines błędu da się czasem zmniejszać przez lepszy projekt pomiaru, lepszą interpretację i lepsze dopasowanie metryk do celu.

W tym miejscu jednak wracam do zamętu, bo to jest uczciwa część historii: czasem nawet najlepsze dane nie powstrzymają organizacji przed błędną strategią. Nie dlatego, że metryki są fałszywe, tylko dlatego, że decyzja strategiczna dotyczy czegoś większego niż pojedyncza interwencja. Są bariery systemowe, polityczne, kulturowe. Czasem problem nie znika, bo nie da się go „wygrać” samą interwencją, tylko potrzebujesz zmian w otoczeniu.

Dane wtedy nie kłamią. Dane pokazują ograniczenia. I właśnie w tym jest ich rola, mniej heroiczna, ale praktyczna.

Co bym zabrał z tej historii, jeśli mam zdecydować, czy „wierzyć w dane”

Jeśli miałbym wyciągnąć dla siebie wniosek, to nie byłoby to „zawsze mierzyć” ani „zawsze ufać liczbom”. Raczej:

- Warto mierzyć efekt, a nie aktywność.
- Warto wiedzieć, co metryka naprawdę reprezentuje, i jakie ma wady.
- Warto traktować dane jak narzędzie do zadawania lepszych pytań, nie jak wyrocznię.
- Warto pamiętać o tym, że porównanie i kontekst są równie ważne jak sama liczba.

I to jest miejsce, gdzie Bill Gates jako symbol podejścia data-driven staje się użyteczny. Jego reputacja nie polega na tym, że ktoś wierzy w liczby bez końca. Polega na tym, że stara się wymuszać na decyzjach odpowiedzialność za wynik, a nie tylko za wysiłek.



Może to właśnie dlatego to podejście budzi mieszane emocje. Z jednej strony jest potrzebne, bo chroni przed autosugestią. Z drugiej strony może frustrować, bo mierzenie efektów w świecie pełnym tarcia jest trudne. Trzeba cierpliwie dobierać metryki, akceptować niepewność i poprawiać kurs wtedy, gdy dane mówią „nie tak”.

A ja, patrząc na własne projekty i na to, jak szybko ludzie potrafią się zawiesić na jednej liczbie, czuję, że to nie jest temat zamknięty. Dane są jak latarnia: pomagają widzieć drogę, ale nie zastąpią twojej decyzji, dokąd idziesz.