ESG a zarządzanie wpływem: od raportowania ESRS do systemów uczenia przyczynowego w świetle teorii Matt Stevensona-Dodda

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje krytyczną różnicę między sprawozdawczością ESG a rzeczywistym zarządzaniem wpływem. Autor wskazuje, że samo raportowanie służy transparentności, podczas gdy zarządzanie wpływem wymaga głębszej analizy przyczynowej i decyzji strategicznych. Kluczowym elementem jest koncepcja podwójnej istotności (double materiality) w standardach ESRS, która rozróżnia wpływ otoczenia na firmę (financial materiality) od oddziaływania firmy na świat (impact materiality). Tekst podkreśla ewolucję wymiaru społecznego ESG, który wykracza poza HR., obejmując cały łańcuch wartości i społeczności lokalne. Według teorii Stevensona-Dodda organizacje muszą odejść od powierzchownych wskaźników na rzecz rygorystycznej ewaluacji rzeczywistych zmian (outcomes), aby uniknąć social washingu i osiągnąć realny wpływ.

The article analyzes the critical difference between ESG reporting and actual impact management. The author points out that reporting itself serves transparency, while impact management requires deeper causal analysis and strategic decisions. A key element is the concept of double materiality in ESRS standards, which distinguishes the environment's impact on the company (financial materiality) from the company's impact on the world (impact materiality). The text emphasizes the evolution of the social dimension of ESG, which extends beyond HR to encompass the entire value chain and local communities. According to Stevenson-Dodd's theory, organizations must move away from superficial indicators toward a rigorous evaluation of actual outcomes to avoid social washing and achieve real impact.

Artykuł stanowi krytyczną analizę przejścia od formalnego raportowania ESG do rzeczywistego zarządzania wpływem społecznym. Autor stawia tezę, że współczesne przedsiębiorstwa często wpadają w pułapkę „compliance sophistication", gdzie perfekcyjne wypełnianie matryc regulacyjnych (np. ESRS) nie idzie w parze z faktyczną wiedzą o skutkach swojej działalności. Tekst argumentuje, że prawdziwa dojrzałość organizacji wymaga budowy wieloczasowej architektury poznawczej: połączenia retrospektywnej ewaluacji przyczynowej, bieżącego monitoringu danych oraz predykcyjnych możliwości AI. Kluczowym wyzwaniem jest tutaj rozróżnienie między korelacją a przyczynowością oraz uniknięcie „automatyzacji błędu" w algorytmach. W konkluzji autor podkreśla, że technologia nie może zastąpić podmiotowości człowieka ani etycznej refleksji nad celem działań. Prawdziwy wpływ (impact) pojawia się dopiero wtedy, gdy wiedza o materialnych skutkach istnienia firmy realnie modyfikuje jej strategię, łańcuch dostaw i kulturę organizacyjną, przekształcając raport ESG z biurokratycznego rytuału w system odpowiedzialności za świat.

SŁOWA KLUCZOWE

ESG | zarządzanie wpływem | raportowanie ESRS | systemy uczenia przyczynowego | teoria Matt Stevensona-Dodda | double materiality | podwójna istotność | impact materiality | financial materiality | causal attribution | CSRD | social washing | limited assurance | Causal ML | value-chain cap

Raportowanie ESG to nie zarządzanie wpływem

ESG jako test dojrzałości zarządzania wpływem: między obowiązkiem ujawnienia, podwójną istotnością a dowodem rzeczywistej zmiany

Polski przypadek prowadzi bezpośrednio do obszaru, w którym termin „wpływ" przeszedł w ostatnich latach wyjątkowo intensywną instytucjonalizację – do sprawozdawczości ESG. To właśnie tutaj najłatwiej jednak popełnić błąd kategorialny. Raportowanie wpływu przedsiębiorstwa, zarządzanie wpływem i naukowa ewaluacja wpływu przyczynowego nie są tym samym procesem, choć posługują się częściowo wspólnym słownictwem. Pierwszy z nich służy transparentności wobec interesariuszy i regulatora, drugi podejmowaniu decyzji wewnątrz organizacji, trzeci zaś

ustaleniu, czy konkretna interwencja rzeczywiście spowodowała obserwowaną zmianę. Pomieszanie tych trzech porządków może prowadzić do paradoksalnej sytuacji, w której przedsiębiorstwo raportuje setki informacji o zrównoważonym rozwoju i posiada formalnie rozwiniętą politykę ESG, a jednocześnie nadal niewiele wie o tym, które z jego działań społecznych zmieniają sytuację ludzi, w jakim stopniu i dlaczego. Stevenson-Dodd przewiduje rosnące znaczenie społecznego wymiaru ESG oraz stopniowe przechodzenie od powierzchownego CSR i wskaźników vanity do bardziej rygorystycznego ujmowania outcomes i impact. Kierunek ten można częściowo potwierdzić, jednak współczesne wydarzenia regulacyjne wymagają znaczącego doprecyzowania tej prognozy.

Rozwój europejskiej sprawozdawczości nie przebiega liniowo – od braku danych ku ich nadmiarowi. W 2026 roku obserwujemy równocześnie dwa procesy: zwiększanie konceptualnej dojrzałości sprawozdawczości wpływu oraz polityczne dążenie do radykalnego zmniejszenia obciążeń raportowych. To napięcie jest niezwykle interesujące z perspektywy całej teorii Stevensona-Dodda.

Jego argument przeciw nadmiarowi wskaźników okazuje się aktualny dokładnie w momencie, gdy europejska regulacja próbuje przejść od ekspansji obowiązków do ich selekcji według materialności. Najważniejszym konceptualnym elementem ESRS jest double materiality, czyli podwójna istotność. Obejmuje ona dwa odrębne, choć powiązane kierunki analizy: impact materiality oraz financial materiality. Pierwszy dotyczy materialnych oddziaływań przedsiębiorstwa na ludzi i środowisko; drugi – materialnych ryzyk i szans związanych ze zrównoważonym rozwojem, które mogą wpływać na sytuację ekonomiczną spółki. Standardy podkreślają przy tym, że wpływ może być materialny społecznie lub środowiskowo niezależnie od tego, czy jest już materialny finansowo. To rozróżnienie ma znaczenie filozoficzne większe, niż sugerowałby język rachunkowości.

Klasyczna analiza strategiczna przedsiębiorstwa patrzy głównie od świata ku firmie: jak regulacje, klimat, rynek pracy, ceny energii albo napięcia społeczne mogą zagrozić rentowności. Impact materiality zmusza do spojrzenia odwrotnego – od firmy ku światu: jak działalność przedsiębiorstwa zmienia warunki pracy, zdrowie, bezpieczeństwo, prawa człowieka, środowisko lub sytuację społeczności.

Można powiedzieć, że podwójna istotność instytucjonalizuje dwa rodzaje przyczynowości. Pierwszy pyta, jak otoczenie oddziałuje na przedsiębiorstwo; drugi – jak przedsiębiorstwo oddziałuje na otoczenie. Ta zmiana jest szczególnie ważna, ponieważ ogranicza redukcję problemów społecznych do tych, które już materializują się w rachunku ekonomicznym. Negatywny wpływ na niewielką społeczność może być materialny z perspektywy impact materiality nawet wtedy, gdy jego

bezpośredni wpływ na wartość przedsiębiorstwa pozostaje niewielki. Tym samym ESRS wprowadzają do sprawozdawczości element, którego czysto finansowa teoria przedsiębiorstwa nie potrafi łatwo pomieścić: znaczenie interesu człowieka nie jest całkowicie uzależnione od jego zdolności wpływania na wynik finansowy spółki. Zarazem materialność nie jest pojęciem równoznacznym z moralnym znaczeniem – jest kategorią selekcyjną sprawozdawczości. Jej zadaniem jest ustalenie, jakie informacje powinny zostać ujawnione, a nie stworzenie kompletnej teorii dobra. Kwestia może być moralnie istotna, a jednocześnie znajdować się poniżej określonego progu materialności raportowej. Dlatego materiality assessment nie zastępuje etyki przedsiębiorstwa; jest sposobem uporządkowania obowiązku informacyjnego.

Rozszerzenie wymiaru społecznego poza granice organizacji

"S" w ESG przestaje oznaczać wyłącznie relacje pracownicze

W zrewidowanej architekturze ESRS wymiar społeczny obejmuje cztery odrębne grupy interesariuszy: własną siłę roboczą, pracowników w łańcuchu wartości, społeczności dotknięte działalnością oraz konsumentów i użytkowników końcowych. Standardy S1-S4 wyznaczają zatem znacznie szerszą sferę społeczną niż klasyczny HR. ESRS S1 dotyczy m.in. warunków pracy, dialogu społecznego, wynagrodzeń, ochrony socjalnej, bezpieczeństwa, rozwoju kompetencji oraz równego traktowania i innych praw pracowniczych; S2 rozszerza analizę na osoby pracujące w łańcuchu wartości, S3 obejmuje społeczności, a S4 konsumentów i użytkowników. Szczególnie istotne jest to, że standard S1 nie ogranicza się do pracowników etatowych, lecz obejmuje także określone kategorie osób niebędących nimi, w tym część samozatrudnionych wykonujących pracę na rzecz przedsiębiorstwa. Standard zastrzega przy tym, że klasyfikacja raportowa nie rozstrzyga ich statusu w prawie pracy.

Jest to kluczowe, gdyż społeczna rzeczywistość produkcji coraz częściej wykracza poza tradycyjną granicę zakładu pracy: przejawia się w outsourcingu, pracy agencyjnej, samozatrudnieniu, podwykonawstwie i globalnych łańcuchach dostaw. Z perspektywy teorii wpływu jest to przejście od organizacyjnego wnętrza do ekosystemu zależności. Przedsiębiorstwo nie może już łatwo ograniczyć analizy społecznej do osób widniejących bezpośrednio na liście płac. Jego strategia zakupowa, warunki kontraktowania czy presja cenowa mogą realnie oddziaływać na ludzi formalnie zatrudnionych gdzie indziej. To kolejny punkt, w którym wpływ społeczny staje się problemem systemowym, a nie tylko programem CSR.

Nowe ESRS są interesujące z jeszcze jednego powodu: nie ograniczają się do prostego ujawniania stanu rzeczy. Struktura standardów społecznych obejmuje polityki, sposoby zaangażowania interesariuszy, działania i zasoby, cele oraz mierniki. W odniesieniu do negatywnego wpływu na pracowników w łańcuchu wartości standard przewiduje również ujawnianie działań służących zapobieganiu, ograniczaniu, zakończeniu lub naprawieniu tego wpływu, a także sposób śledzenia skuteczności tych działań w osiąganiu rezultatów. Konstrukcja ta przypomina uproszczoną Theory of Change: występuje rozpoznany impact, polityka, działanie, zasób, target i próba oceny effectiveness. Nie oznacza to jednak pełnej równoważności z ewaluacją wpływu. Standard sprawozdawczy może wymagać od przedsiębiorstwa odpowiedzi, co robi i jak śledzi skuteczność, ale nie ustanawia automatycznie kontrfaktycznej metodologii pozwalającej wykazać, jaka część zmiany zaszła dzięki konkretnemu działaniu. Właśnie tutaj trzeba postawić ostrą granicę między językiem ESRS i językiem impact evaluation. Jeśli firma wprowadzi nową politykę bezpieczeństwa pracy, a liczba wypadków spadnie, otrzymuje silny sygnał rezultatu. Nie wynika z tego jednak automatycznie, że spadek został spowodowany wyłącznie nową polityką. Mogła zmienić się struktura zatrudnienia, skala produkcji, technologia, sposób rejestracji zdarzeń albo otoczenie regulacyjne.

Rozróżnienie między raportowaniem a rzeczywistym wpływem

ESRS może tworzyć **architekturę odpowiedzialności i śledzenia rezultatów**, lecz pełne causal attribution wymaga dodatkowego aparatu badawczego. To rozróżnienie zasługuje na szczególne podkreślenie, gdyż jego pomijanie jest jednym z głównych źródeł semantycznego zamieszania w dzisiejszym ESG. Słowo **impact** funkcjonuje bowiem w kilku różnych tradycjach intelektualnych.

W ESRS impact oznacza pozytywne lub negatywne oddziaływanie przedsiębiorstwa na ludzi i środowisko. W klasycznej ewaluacji programu odnosi się on częściej do efektu interwencji wyższego rzędu, co wymaga analizy relacji przyczynowej. W impact investing oznacza zarówno zmianę społeczną lub środowiskową, jak i często zamierzoną dodatkowość kapitału. Z kolei w języku marketingowym "impact" bywa synonimem dowolnego pozytywnego działania. Jeśli nie rozdzielimy tych znaczeń, powstanie sytuacja, w której przedsiębiorstwo będzie twierdzić, że "mierzy impact", podczas gdy w rzeczywistości przeprowadza materiality assessment, a organizacja społeczna będzie używać tego samego sformułowania, mając na myśli badanie outcome po

interwencji. Właśnie dlatego naukowa adaptacja Stevenson-Dodda wymaga zbudowania **gramatyki wpływu** również dla ESG. Najbardziej precyzyjny model składałby się z trzech warstw.

ESRS odpowiadałby na pytanie: **jakie oddziaływania są materialne i co przedsiębiorstwo o nich ujawnia**. Impact management odnosiłby się do pytania: **jak zmienia działania i alokację zasobów w odpowiedzi na te informacje**. Impact evaluation natomiast odpowiadałaby na pytanie: **z jaką pewnością możemy stwierdzić, że konkretna interwencja wywołała konkretny rezultat**. Dopiero połączenie tych trzech warstw tworzy przedsiębiorstwo typu "Difference".

Obowiązek raportowania może zwiększyć wiedzę – albo wyprodukować nowy rytuał compliance. Krytyka raportowania sformułowana przez Stevenson-Dodda okazuje się szczególnie użyteczna przy analizie ESG. Każdy nowy wymóg regulacyjny może bowiem wywołać dwa przeciwstawne procesy. W pierwszym wariancie przedsiębiorstwo wykorzystuje obowiązek jako impuls do budowy rzeczywistej infrastruktury danych. Zaczyna mapować łańcuch wartości, zbierać informacje o pracownikach, społecznościach, skargach, wypadkach, nierównościach płacowych i doświadczeniach konsumentów. Dane te zasilają procesy zarządcze, wpływając na procurement, politykę HR, projekt produktu oraz alokację kapitału. W drugim wariancie organizacja buduje przede wszystkim **system produkowania zgodnego raportu**. Tworzy mapę ujawnień, przypisuje właścicieli datapointów, zamawia oprogramowanie, gromadzi dowody i przechodzi assurance, jednak relacja między raportem a decyzją strategiczną pozostaje słaba. Oba przedsiębiorstwa mogą wydać dokument zgodny z ESRS, lecz tylko jedno z nich staje się organizacją uczącą się. CSRD nie eliminuje zatem problemu "Do-Difference" – może go wręcz odtworzyć na wyższym poziomie. Dawniej organizacja typu "Do" liczyła liczbę wolontariuszy.

Compliance nie jest tożsamy z rzeczywistym zarządzaniem wpływem

Organizacja może perfekcyjnie wypełniać matrycę ESRS, pozostając jednocześnie całkowicie niezdolną do modyfikacji swojej działalności podstawowej. **Compliance sophistication nie jest tym samym co impact sophistication.**

Właśnie dlatego wydarzenia z 2026 roku są poznawczo interesujące. 26 lutego 2026 r. opublikowano Dyrektywę Omnibus I (UE 2026/470), która wprowadziła znaczne zawężenie obowiązkowego zakresu CSRD. Według oficjalnych informacji polskiego Ministerstwa Finansów, próg zasadniczy ma teraz obejmować jednostki i grupy przekraczające jednocześnie 1000

pracowników oraz 450 mln euro rocznych przychodów netto. Ograniczono również zakres informacji wymaganych od podmiotów w łańcuchu wartości i pozostawiono atestację na poziomie **limited assurance**, rezygnując z docelowego przejścia na **reasonable assurance**.

Z perspektywy teorii wpływu jest to zmiana kluczowa. Pokazuje ona, że regulator również napotyka problem **kosztu pomiaru**. Im bardziej rozbudowany system danych, tym wyższe koszty transakcyjne, większe ryzyko biurokracji oraz tendencja do przerzucania obowiązków na mniejszych partnerów w łańcuchu wartości. W lipcu 2026 r. Komisja przyjęła dodatkowo zrewidowane standardy ESRS (rozporządzenie delegowane UE 2026/1563). Według Komisji liczba obowiązkowych datapointów spadła o ponad 60%, a całkowita ich liczba o ponad 70%, co ma przełożyć się na obniżenie kosztów raportowania o ponad 30% na przedsiębiorstwo.

Nietrudno tu dostrzec podobieństwo do postulatu Stevensona i Dodda: **więcej danych nie oznacza automatycznie więcej wiedzy**. Regulator zderzył się z tą samą granicą poznawczą, co organizacja społeczna – liczbę ujawnień można zwiększać tylko do momentu, w którym koszt ich produkcji nie zaczyna konkurować z ich wartością informacyjną.

Zawężenie formalnego obowiązku nie oznacza jednak zaniku społecznego wymiaru ESG. Przedsiębiorstwa spoza obowiązkowego zakresu nadal pozostają częścią łańcuchów wartości dużych podmiotów, muszą odpowiadać na wymagania banków, inwestorów czy zamawiających publicznych i mogą dobrowolnie korzystać ze standardów raportowych. Dlatego w 2026 r. Komisja przyjęła dobrowolny standard dla mniejszych firm poza CSRD, wprowadzając mechanizm **value-chain cap**. Ma on ograniczać możliwość przerzucania na małe podmioty żądań informacyjnych wykraczających poza ten standard.

W świetle teorii kosztów transakcyjnych duża firma może obniżyć własny koszt informacyjny, przenosząc ciężar zbierania danych w dół łańcucha. Dla małego dostawcy koszt jednostkowy takiej operacji bywa jednak nieproporcjonalnie wysoki. **Value-chain cap** jest próbą ukrócenia tego zjawiska, czyli **regulatory cost shifting**. Jest to istotne społecznie: system mający zwiększać odpowiedzialność gigantów nie może prowadzić do nadmiernego obciążenia małych dostawców, gdyż w przeciwnym razie sama regulacja wygeneruje negatywny skutek społeczny.

Polska w 2026 r.: system w trakcie transformacji, a nie w stanie ustalonym W polskim porządku prawnym CSRD została wcześniej wdrożona ustawą z 6 grudnia 2024 r., jednak Omnibus I wymusił korektę. W marcu 2026 r. weszła w życie ustawa pozwalająca firmom „pierwszej fali", które po zmianach znalazły się poza zakresem CSRD, na dobrowolne niesporządzanie raportów za lata 2025 i 2026. Pełne merytoryczne wdrożenie Omnibus I pozostaje jednak procesem legislacyjnym. Projekt UC155 ma wprowadzić pozostałe zmiany do ustawy o rachunkowości oraz regulacji

dotyczących atestacji. Według informacji Ministerstwa Finansów z 15 września 2026 r., projekt został przyjęty przez Komitet do Spraw Europejskich i skierowany do Stałego Komitetu Rady Ministrów.

Stan prawny jest zatem w chwili pisania tej sekcji **dynamiczny**. Wymusza to ostrożność wobec uproszczonych tabel typu „kto raportuje od kiedy", gdyż harmonogram krajowej transpozycji wciąż trwa, a termin implementacji dyrektywy unijnej przypada na 19 marca 2027 r. Ta niestabilność ma fundamentalne znaczenie dla **impact management**.

Przedsiębiorstwo budujące system wyłącznie pod aktualny, minimalny obowiązek prawny będzie zmuszone wielokrotnie przebudowywać architekturę danych przy każdej zmianie regulacji. Organizacja projektująca system wokół **materialnych oddziaływań i własnej strategii** wykazuje znacznie większą odporność na zmiany formalne.

„S" jest trudniejsze od „E", ponieważ **człowiek nie jest toną CO_2** Autor książki zakłada, że społeczny wymiar ESG z czasem uzyska rygor zbliżony do wymiaru środowiskowego. Choć kierunek ten jest prawdopodobny, pełna symetria wydaje się metodologicznie wątpliwa. Emisję gazów cieplarnianych można sprowadzić do stabilnej jednostki fizycznej; zużycie energii mierzy się w MWh, a wody w metrach sześciennych. Oczywiście **environmental accounting** również boryka się z problemami granic systemowych czy Scope 3, jednak jego kluczowe zmienne opierają się na mierzalnych jednostkach fizycznych.

Pomiar wpływu społecznego wymaga rygoru przyczynowego i partycypacji

W obszarze „S" przedmiotem analizy są często konstrukty jakościowo odmienne: godna płaca, bezpieczeństwo, wolność zrzeszania się, autonomia, kwestie dyskryminacji, bezpieczeństwo konsumenta, prywatność, dostępność usług czy wpływ na społeczność. Nie istnieje jedna, naturalna jednostka „społecznego oddziaływania", co oznacza, że rozwój pomiaru społecznego impactu nie może polegać wyłącznie na kopiowaniu architektury stosowanej przy emisjach CO_2. Human impact wymaga metod mieszanych: danych jakościowych, walidowanych skal, interpretacji kontekstu oraz aktywnego uczestnictwa osób dotkniętych oddziaływaniem.

Im większa presja na stworzenie jednej liczby w postaci „S-score", tym większe ryzyko utraty treści. Ludzie jako stakeholderzy nie mogą być traktowani wyłącznie jak rekordy w systemie. Zrewidowane ESRS podkreślają wagę zaangażowania grup interesariuszy; w standardach społecznych pojawiają się kanały zgłaszania obaw i potrzeb, dialog oraz monitorowanie

skuteczności działań. Przedsiębiorstwo nie powinno bowiem ustalać społecznego impactu wyłącznie na podstawie własnych danych zarządczych – pracownik, lokalna społeczność czy konsument posiadają wiedzę fenomenologiczną niedostępną centralnemu systemowi danych.

Jednocześnie stakeholder engagement może łatwo stać się jedynie rytuałem konsultacyjnym. Firma może przeprowadzić ankietę, zachowując jednocześnie pełną kontrolę nad tym, które problemy uzna za materialne. Dojrzałość organizacji należałoby więc oceniać nie po liczbie przeprowadzonych konsultacji, lecz przez pryzmat decision rights of evidence: czy informacja od interesariuszy może realnie zmienić materiality assessment, politykę, target, projekt produktu lub warunki zatrudnienia. Jeśli tak nie jest, partycypacja pełni przede wszystkim funkcję legitymizacyjną.

Social washing można zinterpretować jako szczególny przypadek błędu atrybucji. Stevenson-Dodd silnie krytykuje social washing i greenwashing, opisując przejście od publikowania imponujących liczb działań ku audytowalnym rezultatom i wiążąc dojrzałość przedsiębiorstwa z włączeniem wpływu do rdzenia strategii. Z perspektywy wcześniejszej analizy social washing można zdefiniować precyzyjniej: jest to często systematyczny błąd przesuwania się po łańcuchu przyczynowym bez wystarczającego dowodu. Firma finansuje sto szkoleń, więc mówi o „zwiększeniu szans zawodowych". Finansuje program edukacyjny, zatem mówi o „wyrównywaniu nierówności".

Sadzi drzewa, więc mówi o „regeneracji ekosystemów". Wspiera organizację społeczną i włącza jej całkowity rezultat do własnej narracji. W każdym z tych przypadków pominięto pytanie: **So what?**, a bardzo często również to znacznie trudniejsze: "ile z tej zmiany wydarzyłoby się bez nas?"

Social washing nie musi więc wynikać ze świadomego kłamstwa, lecz z metodologicznie słabej konstrukcji przyczynowej. Organizacja może autentycznie wierzyć, że samo finansowanie działania jest równoznaczne ze spowodowaniem impactu. Dlatego najlepszą ochroną przed social washingiem nie jest ostrożniejszy copywriting, lecz lepsza teoria przyczynowości.

Kolejnym punktem potencjalnego nieporozumienia jest assurance. Omnibus I pozostawia obowiązkową atestację sprawozdawczości ESG na poziomie ograniczonego zapewnienia (limited assurance), rezygnując z planowanego przejścia do reasonable assurance. Niezależnie od poziomu zapewnienia należy jednak pamiętać, co assurance rzeczywiście potwierdza: audytor bada jakość i zgodność informacji sprawozdawczej z określonymi kryteriami.

Nie oznacza to automatycznie, że społeczna interwencja była optymalna, sprawiedliwa lub przyczynowo skuteczna. Można posiadać prawidłowo zaatestowany raport opisujący słaby program. Jest to ten sam problem, który pojawił się przy audytowalności outcome: wiarygodność pomiaru

nie jest tożsama z wartością mierzonego działania. Assurance odpowiada na pytanie o wiarygodność sprawozdania. Impact evaluation odpowiada na pytanie o wiarygodność twierdzenia przyczynowego. Ethics pyta, czy cel i sposób działania były właściwe, a Governance – czy organizacja zmieniła decyzję na podstawie wiedzy. Dopiero razem dają one pełną odpowiedzialność.

Wykorzystanie ESRS jako fundamentu dla aktywnego zarządzania wpływem

Najbardziej produktywny sposób połączenia ESRS i koncepcji Stevensona-Dodda polegałby na potraktowaniu ESG nie jako końcowego "dowodu impact", lecz jako **systemu identyfikowania materialnych obszarów wymagających głębszego impact management**. Materiality assessment może wskazać, że szczególnie istotnym problemem jest zdrowie psychiczne pracowników, wypadkowość u podwykonawców, wpływ na lokalną dostępność mieszkań czy dyskryminacja w dostępie do usług.

Na tym etapie proces nie powinien kończyć się raportem. Należy zbudować dla danego problemu Theory of Change: określić mechanizm, działania, rezultaty, wskaźniki, pytania kontrfaktyczne oraz progi decyzyjne. W takim modelu **ESRS identyfikuje materialny problem, a impact management zarządza jego zmianą**. Jest to relacja znacznie dojrzalsza niż mechaniczne utożsamienie obu pojęć; ESG jako architektura uczenia wymaga bowiem danych bliższych operacjom niż te zawarte w raporcie rocznym.

Sprawozdawczość roczna ma naturalnie rytm retrospektywny, co z punktu widzenia accountability jest uzasadnione. Dla zarządzania konkretnymi oddziaływaniami może być jednak zbyt wolna. Wypadki, rotacja, skargi, nierówności, sygnały o naruszeniach w łańcuchu wartości czy rosnące ryzyko dla konsumentów wymagają znacznie krótszych pętli informacji. Dlatego organizacja typu "Difference" potrzebuje **dwóch częstotliwości ESG**. Pierwszą jest częstotliwość regulacyjna, służąca sprawozdawczości. Drugą – częstotliwość zarządcza, dostosowana do dynamiki problemu. Raport roczny może być miejscem syntezy, ale nie powinien być pierwszym momentem, w którym zarząd dowiaduje się o materialnym oddziaływaniu. W praktyce można wyobrazić sobie ewolucję dashboardu ESG w kierunku rozwiązania, które nazwać można **impact cockpit**.

Nie powinien on przedstawiać wszystkich datapointów ESRS jako równorzędnych liczb, lecz budować hierarchię informacyjną. Najwyższy poziom obejmowałby kilka materialnych oddziaływań powiązanych z purpose i strategią przedsiębiorstwa. Drugi pokazywałby outcomes, ich dystrybucję

pomiędzy grupami oraz trend. Trzeci – mechanizmy operacyjne, leading indicators i sygnały ostrzegawcze. Czwarty – dane kontekstowe oraz stopień niepewności przyczynowej. Taka struktura odpowiadałaby Stevenson-Doddowskiemu minimalizmowi znacznie lepiej niż ekran zawierający kilkaset pól raportowych. Regulacja może wymagać wielu ujawnień, ale zarząd potrzebuje **mniejszej liczby lepiej zorganizowanych pytań**. Prawo Goodharta obowiązuje również w ESG.

Gdy określony datapoint zostaje powiązany z ratingiem, wynagrodzeniem menedżera, kosztem finansowania albo dostępem do zamówienia, powstaje bodziec do jego optymalizacji. Może to przyjmować formę pozytywną – gdy przedsiębiorstwo realnie ogranicza negatywny wpływ. Może jednak prowadzić również do zmiany granic raportowania, metodologii czy struktury zatrudnienia, co poprawia wskaźnik bez analogicznej poprawy rzeczywistości społecznej. Jeśli firma mierzy liczbę kobiet w kierownictwie, może poprawić statystykę reprezentacji, nie zmieniając kultury organizacyjnej. Presja na redukcję liczby wypadków w źle zaprojektowanym systemie może prowadzić do ograniczania ich zgłaszania. Z kolei pomiar liczby dostawców "przeszkolonych" z praw człowieka może sprawić, że szkolenie stanie się formalnym substytutem realnego due diligence. Dojrzałe impact management musi więc posiadać **anti-gaming layer**: mierniki jakościowe, dane jakościowe, mechanizmy whistleblowing, independent assurance i analizę outcome, a nie wyłącznie activity compliance.

Jednym z najważniejszych osiągnięć współczesnego ESG jest przesunięcie uwagi z pytania "jakie dobro przedsiębiorstwo robi dodatkowo?" ku pytaniu "jakie skutki tworzy przez własną podstawową działalność?". To filozoficznie istotna zmiana. Firma nie może kompensować naruszeń praw człowieka w łańcuchu dostaw programem stypendialnym; filantropia nie jest prostym offsetem dla szkody społecznej w core business. W języku Theory of Change oznacza to oddzielenie dwóch procesów: **redukcji negatywnego impact i tworzenia dodatkowego pozytywnego impact**. Pierwszy często ma charakter obowiązku moralnego i prawnego, drugi zaś może być strategicznym wyborem przedsiębiorstwa. Nie powinny być one automatycznie kompensowane w jednym indeksie.

To rozróżnienie chroni przed sytuacją, w której firma buduje wysoką "sumę społecznej wartości" poprzez programy charytatywne, podczas gdy największa część jej realnego oddziaływania społecznego powstaje w modelu zatrudnienia, produkcji i sprzedaży. Szansą dla Polski jest wykorzystanie regulacji jako infrastruktury, a nie kolejnego formularza.

W warunkach polskich największa szansa nie polega na tworzeniu przez przedsiębiorstwa objęte CSRD kolejnych równoległych systemów impact measurement. Racjonalniejsze byłoby wykorzystanie infrastruktury zbudowanej dla ESRS jako **wspólnego rdzenia danych**, a

następnie dołączenie warstwy zarządczej odpowiadającej rzeczywistemu purpose przedsiębiorstwa. Dane dotyczące pracowników, dostawców, społeczności i konsumentów mogą zasilać Theory of Change, ale wymagają innej organizacji analitycznej niż sama publikacja raportu. Oznacza to odejście od sytuacji, w której ESG znajduje się wyłącznie w dziale sustainability lub compliance. Jeżeli materialny problem dotyczy pracy w łańcuchu wartości, odpowiedzialność musi przejść do procurement; jeśli konsumentów – do produktu i sprzedaży; jeśli własnej siły roboczej – do HR i zarządu, a jeśli społeczności lokalnej – do inwestycji i operacji. Wtedy impact rzeczywiście staje się strategią.

Stevenson-Dodd miał rację co do kierunku, choć rzeczywistość jest bardziej dialektyczna. Prognoza autorów dotycząca dojrzewania litery "S" jest zatem zasadniczo trafna co do wektora: społeczny wpływ coraz silniej wchodzi do języka zarządów, prawa, inwestorów i assurance. Nie realizuje się on jednak jako prosty marsz ku coraz większej liczbie obowiązkowych danych.

Od raportowania do systemu zarządzania skutkami istnienia

Rok 2026 przyniósł interesujące wnioski. Europejskie regulacje jednocześnie wzmacniają konceptualną strukturę wpływu i redukują zakres raportowania. Zamiast prostej ekspansji mamy próbę znalezienia równowagi między transparentnością a kosztami pozyskania informacji, co można uznać za regulacyjną wersję zasady „measure less, know more" Stevensona-Dodda. Kluczowe nie jest bowiem to, ile punktów danych obejmuje ESRS, lecz czy przedsiębiorstwo dzięki temu systemowi wie więcej o swoich materialnych skutkach i podejmuje inne decyzje.

Z analizy wyłania się szersza koncepcja double accountability. Przedsiębiorstwo odpowiada nie tylko za to, jak świat wpływa na jego wynik, ale również za to, jak jego działania oddziałują na świat. Pierwsza odpowiedzialność jest tradycyjnie silna, gdyż rynek finansowy wymusza jej przestrzeganie – zarząd ignorujący zagrożenia dla rentowności szybko ponosi konsekwencje. Druga pozostaje historycznie słabsza, ponieważ koszty oddziaływania można eksternalizować na pracowników, społeczności, państwo lub przyszłe pokolenia.

Double materiality jest próbą przywrócenia symetrii epistemicznej: przedsiębiorstwo ma dostrzec oba kierunki zależności. Stevenson-Dodd dodaje do tej konstrukcji element, którego samo raportowanie nie zapewni: obowiązek przełożenia wiedzy na zmianę działania. Właśnie dlatego najbardziej dojrzałą formą ESG nie będzie doskonały raport, lecz organizacja zdolna wykazać, że rozpoznanie materialnego negatywnego wpływu doprowadziło do zmiany w obszarze

procurementu, produktu, polityki zatrudnienia, inwestycji lub warunków współpracy – a następnie, że skuteczność tej zmiany została zweryfikowana. Dopiero wtedy litera „S” przestaje być rozdziałem raportu i staje się systemem zarządzania skutkami własnego istnienia.

Warto przejść od regulacyjnej teraźniejszości ku przyszłości mierzenia wpływu: modelowi Trzech Er Stevensona-Dodda, AI, Big Data, otwartym danym i predykcji potrzeb społecznych. W tym obszarze należy oddzielić uzasadnioną technologicznie zmianę od futurologicznego entuzjazmu oraz zbadać pytanie: czy system, który dzięki AI potrafi przewidywać potrzeby człowieka bez zadawania mu pytań, rzeczywiście rozumie go lepiej, czy jedynie sprawniej oblicza jego cyfrowe ślady.

Stevenson-Dodd opisuje ewolucję pomiaru wpływu poprzez model Trzech Er. Pierwszą jest retrospektywna ewaluacja wykonywana po zakończeniu programu; drugą – środowisko Live Data, obejmujące chmury, CRM, API i dashboardy; trzecią zaś integracja rozległych zbiorów danych z uczeniem maszynowym i sztuczną inteligencją, co ma umożliwiać nie tylko obserwację rezultatów, lecz prognozowanie potrzeb i kierowanie interwencji, zanim problem osiągnie pełną skalę. Autorowi towarzyszy tu idea bezankietyzacji: wykorzystania zastanych danych administracyjnych, środowiskowych i operacyjnych zamiast wielokrotnego angażowania beneficjentów w formularze badawcze. Konstrukcja ta jest heurystycznie wartościowa, jednak jako periodyzacja rozwoju nauki wymaga doprecyzowania.

Kolejne ery nie zastępują poprzednich, lecz nakładają się na siebie, rozwiązując różne problemy poznawcze. Retrospektywna ewaluacja nie staje się metodologicznie przestarzała przez pojawienie się dashboardów. Live Data nie zastępuje kontrfaktyczności, a sztuczna inteligencja nie zamienia korelacji w przyczynowość. Każdy kolejny poziom przede wszystkim skraca opóźnienie informacyjne, zwiększa skalę przetwarzania lub rozszerza zdolność wykrywania wzorców, co nie gwarantuje automatycznie silniejszego dowodu o wpływie.

Trafniejszy od modelu trzech następujących po sobie epok byłby zatem model trzech równoległych funkcji systemu wiedzy: ewaluacja retrospektywna służy identyfikacji skutków i głębokiej refleksji; Live Data – bieżącej regulacji działania; AI i Big Data – wykrywaniu wzorców, prognozowaniu, segmentacji oraz wspomaganiu decyzji. Dojrzała organizacja przyszłości będzie prawdopodobnie potrzebowała wszystkich trzech funkcji jednocześnie.

Wieloczasowa architektura poznawcza: od detekcji do wyjaśniania

Autor książki przedstawia ewaluację retrospektywną w sposób dość polemiczny, ukazując ją jako kosztowną, opóźnioną i słabo powiązaną z bieżącym zarządzaniem. Krytyka ta jest zasadna wtedy, gdy wielomiesięczne badanie służy wyłącznie formalnemu zamknięciu projektu, a jego rekomendacje pojawiają się zbyt późno, by wpłynąć na decyzje. Nie można jednak utożsamiać opóźnienia z brakiem wartości epistemicznej. Niektórych rezultatów nie da się wiarygodnie zmierzyć szybko, ponieważ sama ich definicja zawiera wymiar czasu. Trwałość zatrudnienia poznajemy dopiero po upływie miesięcy; recydywę, efekty zdrowotne, mobilność edukacyjną, rozwój dziecka czy trwałość zmiany społecznej często trzeba obserwować latami. Natychmiastowy dashboard może zatem dostarczać doskonałej wiedzy o tym, co wydarzyło się dzisiaj, lecz bardzo słabej informacji o tym, co okaże się istotne za trzy lata. Szybkość i jakość informacji nie pozostają więc w relacji monotonnej. Niektóre pytania wymagają krótkiej pętli zwrotnej, inne zaś cierpliwości poznawczej. Jeszcze ważniejsza jest jednak kwestia przyczynowości.

Randomizowane eksperymenty, badania quasi-eksperymentalne, difference-in-differences, regression discontinuity czy synthetic control wymagają często stabilnej definicji interwencji, obserwacji odpowiedniego horyzontu oraz starannego konstruowania grup porównawczych. To właśnie wolniejsza analiza może dostarczyć wiedzy niedostępnej dla systemów Live Data: odpowiedzi na pytanie, czy dany rezultat wystąpił rzeczywiście dzięki programowi. Dlatego przyszłość nie powinna polegać na zastąpieniu ewaluacji dashboardem, lecz na przejściu od przeciwstawiania obu metod ku wieloczasowej architekturze poznawczej. Organizacja potrzebuje szybkiego monitoringu do bieżącej korekty działań oraz wolniejszej ewaluacji, by sprawdzić, czy samo działanie ma przyczynowy sens. Druga era wnosi tu rzeczywistą zmianę organizacyjną.

Dane, które wcześniej trafiały do zarządu po roku, mogą być teraz obserwowane tygodniowo, dziennie, a w niektórych przypadkach niemal na bieżąco. Autor opisuje tę transformację poprzez integrację CRM, chmury, API oraz dashboardów, wskazując na możliwość śledzenia ścieżki beneficjenta od rekrutacji aż po długotrwały rezultat. Z metodologicznego punktu widzenia kluczowym skutkiem jest zmniejszenie information latency, czyli opóźnienia pomiędzy zmianą w rzeczywistości a dostępnością informacji o tej zmianie. Jeśli odpływ uczestników wzrasta w marcu, organizacja nie musi odkrywać tego w raporcie sporządzonym następnej zimy; może zidentyfikować problem wystarczająco wcześnie, by zbadać jego przyczynę. Nie zmienia się jednak logiczny status obserwacji. Jeżeli wskaźnik zatrudnienia spada, dashboard informuje o spadku, ale nie odpowiada na pytanie „dlaczego”. Jeszcze mniej mówi o tym, czy zmiana wynika z jakości

programu, sezonowości, koniunktury, selekcji uczestników, zmiany struktury populacji czy czynnika zewnętrznego. Live Data radykalnie zwiększa zatem zdolność detekcji, lecz zdolność wyjaśniania wciąż wymaga Theory of Change, statystyki, badań jakościowych i analizy przyczynowej. Jest to różnica analogiczna do tej, którą medycyna dostrzega między monitorem pokazującym spadek saturacji a diagnozą wyjaśniającą przyczyny tego spadku.

AI i Big Data wprowadzają do tego układu funkcję jakościowo nową: możliwość przewidywania. Model może uczyć się wzorca P(Y|X), czyli prawdopodobieństwa wystąpienia określonego zdarzenia przy znanych cechach X. Pozwala to prognozować ryzyko przerwania programu, bezrobocia, zaległości czynszowej, hospitalizacji czy niedostatecznego wykorzystania usługi. Autor idzie w tym kierunku bardzo daleko, prezentując połączenie informacji o efektywności energetycznej budynków, deprywacji i zużyciu energii jako przykład identyfikacji gospodarstw zagrożonych ubóstwem energetycznym bez prowadzenia klasycznej ankiety. Choć sam kierunek analityczny jest wiarygodny, szczegółowe twierdzenia autora o dokładności tego konkretnego rozwiązania nie powinny być traktowane jako zweryfikowany wynik naukowy bez stosownej dokumentacji metodologicznej. Przykład ten dobrze ilustruje jednak coś ważniejszego: przejście od danych zebranych specjalnie dla ewaluacji do danych powstających niezależnie od niej. To fundamentalna zmiana; tradycyjny badacz konstruował kwestionariusz, wybierał próbę i pozyskiwał odpowiedzi.

W środowisku Big Data część obserwacji istnieje wcześniej jako produkt funkcjonowania systemu: są to rejestry administracyjne, dane o transakcjach, mobilności, energii, edukacji, usługach publicznych czy użytkowaniu systemów cyfrowych. Ewaluator może zatem działać nie tylko jako producent danych, lecz jako integrator istniejących śladów instytucjonalnych. Bank Światowy w Government Analytics Handbook ujmuje administracyjne zbiory danych właśnie jako jeden z podstawowych zasobów nowoczesnej analityki państwa i łączy ich wykorzystanie z zasadą „measuring what matters", problem-solving oraz uczeniem organizacyjnym. W Unii Europejskiej podobną logikę infrastrukturalną rozwijają wspólne europejskie przestrzenie danych, które mają pozwalać na bezpieczne współdzielenie informacji w obszarach takich jak zdrowie, energia czy administracja publiczna. Komisja akcentuje przy tym wspólne mechanizmy governance, interoperacyjność i ochronę prywatności. Przyszłość pomiaru wpływu może więc zależeć mniej od istnienia jednego gigantycznego zbioru, a bardziej od zdolności bezpiecznego łączenia wielu rozproszonych źródeł.

Różnica między predykcją a wnioskowaniem przyczynowym

Kluczowa granica możliwości AI przebiega jednak między **prediction i intervention**. Model może z dużą precyzją przewidzieć, kto prawdopodobnie pozostanie bezrobotny, a jednocześnie nie wiedzieć, jaka konkretna interwencja poprawi sytuację danej osoby. To fundamentalna różnica między pytaniem „co prawdopodobnie się wydarzy?" a pytaniem „co się stanie, jeśli zrobimy X zamiast Y?". Pierwsze z nich jest domeną analityki predykcyjnej, drugie zaś – wnioskowania przyczynowego.

Jeśli model wykaże silny związek między samotnym rodzicielstwem a długotrwałym bezrobociem, nie oznacza to automatycznie, że interwencja skierowana na dowolną cechę skorelowaną z tą sytuacją poprawi szanse na zatrudnienie. Kluczowym mechanizmem może być bowiem brak opieki nad dziećmi, lokalizacja, struktura czasu pracy, stan zdrowia lub specyficzne normy pracodawców.

Ryzyka automatyzacji i pułapki optymalizacji algorytmicznej

Model predykcyjny opisuje ryzyko, jednak strategia wymaga dźwigni przyczynowej. Współczesny rozwój **causal machine learning** dąży właśnie do połączenia bogactwa metod ML z teorią przyczynowości i analizą potencjalnych wyników. Szczególnie istotnym kierunkiem jest estymacja **heterogeneous treatment effects**: zamiast pytać wyłącznie o średni efekt programu, próbuje się określić, jak różni się on w zależności od cech poszczególnych osób. Przegląd opublikowany w **Annual Review of Biomedical Data Science** w 2025 r. wskazuje na rozwój metod łączących machine learning z **causal inference**, co ma umożliwić przejście „od predykcji do preskrypcji". Dla impact management oznacza to istotną zmianę: organizacja przestaje pytać jedynie, czy program A jest średnio skuteczniejszy od programu B, a zaczyna badać, dla kogo A wypada lepiej niż B. To nie jest drobna korekta statystyczna, lecz przejście od jednolitego programu dla „przeciętnego beneficjenta" do heterogenicznej teorii interwencji. W tym miejscu pojawia się jednak poważny problem normatywny.

Algorytm optymalizujący oczekiwany efekt może racjonalnie kierować zasoby do osób o największej **responsiveness**, czyli najwyższym przewidywanym przyroście outcome wskutek interwencji. Pozwala to maksymalizować sumę korzyści, ale jednocześnie prowadzi do systematycznego

pomijania osób najtrudniejszych. Powraca zatem problem **creaming**, tym razem w technologicznie znacznie potężniejszej postaci.

Jeżeli model oceni, że osoba A ma 80-procentową szansę poprawy, a osoba B jedynie 15-procentową, system nastawiony na maksymalizację liczby sukcesów zacznie preferować osobę A. Jeśli jednak to B znajduje się w znacznie gorszym położeniu, taka polityka może naruszać zasadę pierwszeństwa potrzeb. AI nie rozwiązuje konfliktu między efektywnością a sprawiedliwością; może go jedynie zautomatyzować i ukryć w funkcji celu. Właśnie dlatego optymalizacja algorytmiczna musi ujawniać, co właściwie optymalizuje. Funkcja celu jest zawsze teorią wartości zapisaną w formie matematycznej. Maksymalizacja **employment outcome**, QALY, wellbeing czy liczby obsłużonych osób to odmienne decyzje moralne, nawet jeśli wszystkie wykonuje ten sam algorytm.

Kolejny problem wynika z genezy danych. Model uczący się na danych historycznych przejmuje nie tylko obraz rzeczywistości, lecz także historyczne praktyki instytucjonalne. Jeśli określona grupa miała utrudniony dostęp do programu, dane będą zawierać mniej przykładów jej sukcesu. Jeśli pewne zachowania były częściej kontrolowane, odnotowano więcej naruszeń. Jeśli konkretne społeczności rzadziej korzystały z usług, baza może sugerować mniejszy popyt, choć realną barierą był brak dostępu.

Machine learning może wówczas utrwalić zależność historyczną i przedstawić ją jako predykcyjną prawidłowość. Jest to szczególnie ryzykowne w sektorze publicznym, gdzie predykcja sama zmienia rozkład zasobów: model przewiduje mniejsze prawdopodobieństwo powodzenia u danej grupy, więc grupa otrzymuje mniej wsparcia; przez brak wsparcia osiąga gorsze wyniki, co z kolei potwierdza wcześniejszą predykcję. Powstaje samowzmacniająca pętla klasyfikacyjna.

Dlatego **fairness** nie może być oceniana wyłącznie przez pryzmat **accuracy** całego modelu. Niezbędna jest analiza błędów między grupami, badanie możliwości samorealizacji prognozy oraz weryfikacja tego, jakie dane pozostają nieobserwowane. W tym punkcie impact measurement styka się z problemem performatywności. Tradycyjna statystyka często zakłada, że obserwator nie zmienia istotnie badanego systemu. W systemach społecznych założenie to jest kruche, a algorytmiczne zarządzanie czyni je jeszcze bardziej problematycznym. Jeśli model zacznie kierować interwencje do wybranych osób, zmieni ich przyszłe rezultaty. Gdy wyniki te zostaną wykorzystane do trenowania kolejnej wersji modelu, dane nie opisują już świata „bez algorytmu", lecz świat częściowo przez niego ukształtowany. System predykcyjny staje się elementem procesu, który sam przewiduje.

W konsekwencji przyszły impact management musi być coraz bardziej świadomy **policy feedback loops** i performatywności danych. Model powinien być okresowo rekalibrowany, a jego wpływ na

strukturę populacji i zachowania pracowników badany tak samo, jak skutki wspieranego przez niego programu.

OECD w raporcie **Governing with Artificial Intelligence** z 2025 r. ostrzega przed **automation bias**, nieprzejrzystością modeli oraz ryzykiem utrwalania błędów na dużą skalę. Zwraca uwagę, że pracownicy administracji mogą nadmiernie ufać wynikom systemu, ponieważ wyglądają one jak bezstronna kalkulacja, podczas gdy błąd pozostaje trudny do zaobserwowania. Jest to istotne uzupełnienie optymistycznej wersji „Ery 3".

Automatyzacja nie skaluje jedynie kompetencji. Skaluje wszystko, co zostało zakodowane – również błąd, zły proxy, nieaktualną teorię, historyczne uprzedzenie i niewłaściwą funkcję celu. Jeśli analityk pomyli się ręcznie przy stu sprawach, szkoda jest ograniczona. Jeżeli jednak błędny model zostanie włączony do infrastruktury obsługującej milion osób, błąd uzyskuje przemysłową skalę.

AI i automatyzacja danych nie mogą zastąpić podmiotowości człowieka

Rozwój AI nie zmniejsza zatem znaczenia governance – wręcz przeciwnie, zwiększa je. OECD w raporcie **Digital Government Outlook 2026** wskazuje, że transparentność, algorithmic accountability oraz impact assessment pozostają jednymi z najsłabiej rozwiniętych obszarów cyfrowego państwa w krajach członkowskich, mimo gwałtownej ekspansji zastosowań AI. Oznacza to, że technologiczna dojrzałość implementacji może wyprzedzać dojrzałość instytucji zdolnych ją kontrolować. W tym kontekście Europejski AI Act wyznacza wyraźną granicę między pomocą w analizie a algorytmicznym władaniem człowiekiem.

W sektorze społecznym kluczowe jest ryzyko, że AI przestanie być narzędziem opisującym potrzeby, a stanie się elementem decyzji o dostępie do praw i zasobów. Europejski AI Act traktuje ten obszar jako szczególnie wrażliwy. Systemy wykorzystywane przez organy publiczne do oceny uprawnień do podstawowych świadczeń i usług publicznych – w tym części świadczeń zdrowotnych, społecznych i mieszkaniowych – zakwalifikowano do kategorii wysokiego ryzyka. Rozporządzenie wprost wskazuje na możliwość naruszenia ochrony socjalnej, zasady niedyskryminacji, godności oraz prawa do skutecznego środka prawnego. Jednocześnie AI Act zabrania określonych form social scoringu, w których ocena zachowania lub przewidywanych cech człowieka prowadzi do nieuzasadnionego lub nieproporcjonalnego niekorzystnego traktowania.

Zgodnie z aktualnym harmonogramem, po zmianach przyjętych w 2026 r., szczegółowe obowiązki dotyczące przypadków wysokiego ryzyka (wymienionych w załączniku III) zaczną obowiązywać od

2 grudnia 2027 r., podczas gdy część pozostałych przepisów aktu jest już stosowana. Z punktu widzenia teorii wpływu ta granica jest zasadnicza. Można wykorzystać AI, aby zidentyfikować obszar wymagający większej uwagi, jednak znacznie poważniejszą decyzją jest automatyczne uznanie konkretnej osoby za mniej wartą wsparcia. Predykcja ryzyka i przyznawanie prawa to dwa odmienne akty epistemiczne i polityczne.

„Bezankietyzacja" może zmniejszyć obciążenie, ale jednocześnie pogłębić epistemiczne milczenie. Autor książki traktuje ograniczenie ankiet jako jedną z głównych korzyści Trzeciej Ery. Jest to argument przekonujący tam, gdzie organizacje wielokrotnie pytają beneficjentów o informacje już dostępne w systemach administracyjnych; nie ma powodu, by człowiek po raz dziesiąty podawał dane, które instytucja może legalnie i bezpiecznie pozyskać w inny sposób. Jednocześnie „bezankietyzacja" niesie ze sobą epistemiczną cenę. Dane administracyjne rejestrują bowiem przede wszystkim to, co instytucja wcześniej zdecydowała się ewidencjonować. Są one doskonałe do ustalenia, czy ktoś otrzymuje świadczenie, płaci składkę, pozostaje zatrudniony lub korzysta z usługi, lecz znacznie słabiej oddają poczucie godności, doświadczenie dyskryminacji, jakość relacji, aspiracje czy powody rezygnacji. Można zatem stworzyć system, który wie o człowieku bardzo dużo i nigdy go o nic nie pyta – technologicznie będzie to imponujące, epistemologicznie zaś może stanowić regres. Dojrzałą zasadą nie powinno być zatem „zero ankiet", lecz zero zbędnych pytań i zero utraconego głosu. Informacja zastana powinna zastępować respondentowi pracę administracyjną, ale nie jego prawo do opowiedzenia własnego doświadczenia. Łączenie danych przestrzennych rodzi w tym zakresie dodatkową trudność.

Epistemologia stosowana zamiast ślepej wiary w dane

Jeżeli dzielnica charakteryzuje się wysokim poziomem deprywacji, nie oznacza to, że każda osoba mieszkająca na jej obszarze znajduje się w trudnej sytuacji. Jest to klasyczny **ecological fallacy**: nieuprawnione przenoszenie zależności zaobserwowanej na poziomie agregatu na poziom jednostkowy. Właśnie dlatego dane obszarowe mogą być świetnym narzędziem alokacji działań terenowych, lecz słabszym instrumentem indywidualnego orzekania o potrzebach. Model może wskazać kwartał miasta, w którym prawdopodobieństwo ubóstwa energetycznego jest wysokie i skierować tam mobilny zespół, ale nie powinien – bez dodatkowej weryfikacji – rozstrzygać, że konkretne gospodarstwo jest ubogie tylko dlatego, że znajduje się w strefie wysokiego ryzyka. Rozróżnienie pomiędzy **area targeting** a **individual adjudication** powinno stać się jednym z fundamentów przyszłej etyki analityki społecznej.

Wielkie zbiory danych nie oznaczają pełnych danych Pojęcie Big Data wytwarza czasem złudzenie epistemicznej kompletności. Przy milionach rekordów w bazie łatwo zapomnieć, że system nadal może systematycznie pomijać pewne fragmenty rzeczywistości. Wielkość próby redukuje błąd losowy, ale nie eliminuje biasu selekcyjnego. Jeśli osoby w najgłębszym kryzysie zostawiają mniej cyfrowych śladów, mają niestabilny adres, brak rachunku bankowego lub rzadziej korzystają z formalnych usług, stają się w systemach Big Data niemal niewidoczne. Brak danych nie jest wówczas przypadkowy – jest społecznie ustrukturyzowany.

W konsekwencji system najbardziej „data-rich" może być jednocześnie ślepy na osoby **data-poor**. To zasadniczy problem analityki społecznej, gdyż populacje najbardziej potrzebujące wsparcia bywają najsłabiej reprezentowane w stabilnych rejestrach. Najbardziej obiecującą ścieżką rozwoju nie jest więc samo Big Data, lecz połączenie uczenia maszynowego z **causal inference**. Klasyczny model predykcyjny pyta, kto prawdopodobnie osiągnie dany wynik. Model causal ML próbuje natomiast szacować, u kogo konkretna interwencja ten wynik zmieni.

Osoba z bardzo wysokim ryzykiem bezrobocia może zachować to samo ryzyko nawet po standardowej interwencji – program nie zmienia wyniku, mimo że wskaźnik był wysoki. Inna osoba może mieć średnie ryzyko bazowe, ale silnie reagować na określony typ wsparcia. Jeśli celem jest realny wpływ (**impact**), ważniejsza od samego prawdopodobieństwa problemu staje się różnica między potencjalnymi wynikami z interwencją i bez niej. W tym miejscu technologia powraca do kontrfaktycznego rdzenia teorii wpływu z części IV. AI przyszłości może okazać się wartościowa nie dzięki „większej inteligencji", lecz możliwości operacyjnego szacowania heterogenicznej dodatkowości interwencji.

Warunkiem pozostaje jednak poprawna identyfikacja przyczynowa. Żaden model machine learning nie eliminuje confoundingu samą mocą obliczeniową. Jeśli dane nie pozwalają odróżnić leczenia od selekcji do leczenia, algorytm może jedynie dokładniej estymować błędną zależność. Autor książki przewiduje, że kluczową kompetencją liderów stanie się **prompt engineering**, czyli zdolność formułowania precyzyjnych zapytań do systemów AI. Prognozę tę warto dziś skorygować: umiejętność komunikacji z modelami pozostanie użyteczna, lecz interfejsy będą prawdopodobnie upraszczane, a samo konstruowanie promptów automatyzowane. Trwalszą kompetencją będzie coś znacznie głębszego: **problem formulation**. Lider musi rozumieć, jakie pytanie naprawdę chce rozstrzygnąć; czy jest ono opisowe, predykcyjne, przyczynowe czy normatywne; jakie dane mogą dać odpowiedź; które zmienne są skutkiem, a które przyczyną lub confounderem; który błąd jest kosztowniejszy i kogo dane nie reprezentują. Osoba nieumiejąca odróżnić predykcji od przyczynowości może stworzyć technicznie doskonały prompt i otrzymać precyzyjną odpowiedź na

źle postawione pytanie. Dlatego przyszła kompetencja lidera impact przypomina bardziej epistemologię stosowaną niż sztukę pisania instrukcji dla modelu.

OECD w analizach AI w administracji akcentuje znaczenie governance, danych, kompetencji, infrastruktury, inwestycji, procurement i partnerstw jako równoległych warunków wiarygodnego wykorzystania technologii. Narzędzia bez instytucjonalnej zdolności rozumienia ich ograniczeń nie stworzą inteligentnego państwa ani organizacji. Autor książki idzie dalej, prognozując nadejście AGI zdolnej do rozwiązywania wielowymiarowych problemów społecznych. W analizie należy jednak wyraźnie oddzielić ten fragment od twierdzeń empirycznie ugruntowanych. Nie istnieje obecnie podstawa do traktowania przyszłego pojawienia się AGI, jego terminu ani zdolności samodzielnego „rozwiązywania problemów społecznych" jako ustalonego faktu.

Co więcej, nawet hipotetycznie zaawansowany system poznawczy nie usuwa normatywnego charakteru problemów społecznych. Bezdomność, nierówność czy migracja nie są jedynie problemami optymalizacyjnymi o jednej funkcji celu; zawierają konflikty interesów i teorii sprawiedliwości. AI może coraz lepiej modelować konsekwencje decyzji, ale nie oznacza to, że może obiektywnie ustalić, która wizja sprawiedliwego społeczeństwa powinna zostać wybrana. Właśnie ta różnica wyznacza granicę pomiędzy inteligencją obliczeniową a legitymacją polityczną.

Demokratyzacja AI wymaga infrastruktury etycznej i ochrony relacji

Autor przewiduje, że spadek kosztów narzędzi AI umożliwi małym organizacjom i lokalnym społecznościom korzystanie z zaawansowanej analityki, która dotychczas była dostępna głównie dużym instytucjom. Choć scenariusz ten jest realny, demokratyzacji narzędzi może towarzyszyć koncentracja infrastruktury. Sam model może być tani w użyciu, jednak dane wysokiej jakości, moc obliczeniowa, kompetencje, cyberbezpieczeństwo, integracje oraz legalny dostęp do rejestrów nie są rozłożone równomiernie. Mała organizacja może korzystać z tego samego modelu językowego co korporacja, ale brakować jej będzie zespołu data science, prawnika od ochrony danych czy infrastruktury do interoperacyjnego łączenia rejestrów.

Technologia może więc równocześnie obniżać próg wejścia i zwiększać premię za zdolności komplementarne. W ekonomii technologii jest to klasyczny problem skill-biased technological change: narzędzie staje się powszechne, lecz najwyższą produktywność osiągają ci, którzy dysponują odpowiednimi kompetencjami i zapleczem. Dlatego prawdziwa demokratyzacja impact analytics wymaga nie tylko taniego AI, ale także otwartych standardów, publicznej infrastruktury

danych, edukacji metodologicznej, wspólnych narzędzi oraz finansowania capacity building. Jednym z najbardziej prawdopodobnych kierunków jest odejście od wizji jednej, gigantycznej bazy danych. Wspólne europejskie przestrzenie danych projektuje się raczej jako infrastruktury umożliwiające kontrolowany dostęp i przetwarzanie informacji przy zachowaniu reguł governance oraz praw właścicieli danych. Z punktu widzenia analizy wpływu otwiera to interesującą możliwość: dane mogą być analizowane wspólnie bez konieczności ich pełnego scentralizowania. Rozwijające się techniki federated learning, privacy-preserving computation, trusted research environments czy differential privacy wskazują kierunek, w którym organizacje mogą uzyskiwać wiedzę o populacji bez kopiowania wszystkich danych osobowych do jednego repozytorium. Jest to szczególnie istotne dla administracji publicznej i sektora społecznego.

Im bardziej szczegółowy jest model człowieka, tym większa jego wartość predykcyjna, ale i ryzyko naruszenia prywatności. Przyszłość nie powinna więc być opisana prostym równaniem więcej danych = lepsza wiedza. Trafniejsze brzmienie to: lepsze zarządzanie dostępem + odpowiednia metoda + minimalizacja danych + właściwy cel = większa wartość informacyjna przy akceptowalnym ryzyku. Jedna z najcenniejszych prognoz autora dotyczy paradoksalnego wzrostu wartości kompetencji relacyjnych wraz ze spadkiem kosztów analizy i generowania informacji. Stevenson-Dodd wskazuje na empatię, słuchanie, relację i zaufanie jako zasoby, których znaczenie będzie rosło w coraz bardziej zautomatyzowanym środowisku. Nie należy interpretować tego jako nostalgicznej opozycji „człowiek kontra maszyna". W wielu usługach relacja jest częścią mechanizmu przyczynowego.

Zaufanie do terapeuty może wpływać na adherencję, relacja z mentorem na poczucie własnej skuteczności, a bezpieczna więź z pracownikiem socjalnym na ujawnienie informacji niezbędnych do właściwej pomocy. Jeżeli organizacja automatyzuje element będący nośnikiem mechanizmu, może zwiększyć efektywność (efficiency), ale jednocześnie zmniejszyć realny wpływ (impact). Dlatego kluczowe pytanie nie brzmi, które zadania „AI potrafi wykonać", lecz które zadania można zautomatyzować bez uszkodzenia mechanizmu zmiany. Jest to znacznie wyższy standard. Paradoks ten będzie szczególnie widoczny w usługach społecznych. Organizacja może automatycznie transkrybować rozmowy, analizować sentyment, klasyfikować potrzeby i sugerować interwencje.

Dzięki temu może uwolnić czas, ale może również doprowadzić do sytuacji, w której każda rozmowa staje się źródłem danych, a człowiek zaczyna doświadczać relacji pomocowej jako procesu ciągłego profilowania. Sama możliwość pomiaru potrafi zmienić zachowanie, co prowadzi do koncepcji measurement reactivity. Beneficjent nie zawsze mówi to samo w intymnej rozmowie, a w sytuacji, gdy wie, że każda jego wypowiedź zostanie automatycznie zakodowana.

Od analityki danych do aksjologii odpowiedzialności

Technologia mierząca wpływ może wpływać na to, co mierzy. Właśnie dlatego systemy przyszłości potrzebują nie tylko dokładności i prywatności, ale także **relational legitimacy** – społecznej akceptowalności sposobu poznawania człowieka. Im bliżej decyzji o zatrudnieniu, świadczeniu, kredycie, edukacji, zdrowiu czy pomocy społecznej znajduje się system pomiarowy, tym mniej jest on wyłącznie narzędziem analitycznym, a staje się częścią infrastruktury władzy. Klasyfikacja rozdziela ludzi na grupy, ranking ustala kolejność, a predykcja ryzyka kieruje uwagę instytucji.

Systemy rekomendacyjne sugerują wybory, zaś automatyczne decyzje mogą przyznać lub odmówić zasobów. W takim świecie pytania o wyjaśnialność (explainability), możliwość zakwestionowania decyzji, niedyskryminację i kontrolę człowieka nie są jedynie dodatkiem w ramach „ethical AI", lecz stają się częścią konstytucyjnej architektury relacji między jednostką a instytucją. To właśnie dlatego europejskie regulacje szczególnie rygorystycznie traktują wykorzystanie AI przy świadczeniach publicznych i prawach podstawowych. Przyszłość pomiaru wpływu nie będzie więc kształtowana wyłącznie przez data scientists, lecz współtworzona przez prawników, filozofów, ekonomistów, psychologów, socjologów, specjalistów od governance, organizacje obywatelskie oraz samych ludzi poddawanych klasyfikacji.

W całej dyskusji o cyfrowej przyszłości szczególnie niefortunne wydaje się określenie **data-driven**. Sugeruje ono, że dane mogą prowadzić organizację, tymczasem dane są produktem kategorii, procesów rejestracji, selekcji i modeli. Zawierają w sobie historię instytucji, która je wytworzyła.

Dojrzalszym ideałem byłaby organizacja **causally informed, evidence-responsive and rights-constrained**: taka, która wykorzystuje dane do aktualizowania hipotez, rozróżnia predykcję od przyczynowości i jest zdolna testować heterogeniczne efekty, pozostając jednocześnie ograniczoną prawami człowieka oraz jawnie określoną aksjologią. W takim modelu AI nie podejmuje „obiektywnej decyzji", lecz dostarcza dodatkowego typu dowodu w procesie decyzyjnym. Im większe są konsekwencje decyzji dla życia człowieka, tym bardziej powinna wzrastać nie tylko dokładność modelu, lecz również **liczba instytucjonalnych zabezpieczeń przed błędem modelu**.

Jeżeli model Trzech Er potraktować jako punkt wyjścia, można zaproponować jego dalszą ewolucję. Czwarty etap nie musiałby oznaczać kolejnego wzrostu mocy obliczeniowej, lecz **integrację monitoringu, ewaluacji przyczynowej, danych jakościowych, AI i deliberacji w jeden adaptacyjny system uczenia**. W takim układzie dashboard wskazuje anomalię, model predykcyjny identyfikuje grupę ryzyka, a Causal ML pomaga ocenić, która interwencja może przynieść największy efekt dla konkretnej grupy. Dane jakościowe ujawniają mechanizm i

doświadczenie, a Theory of Change interpretuje całość, podczas gdy reguły prawne ograniczają dopuszczalne interwencje.

Lived experience weryfikuje, czy wynik odpowiada rzeczywistemu rozumieniu dobra. Zarząd podejmuje decyzję, a kolejne dane badają jej konsekwencje. Nie jest to już „pomiar" w tradycyjnym sensie, lecz **ciągły system uczenia przyczynowego i normatywnego**. Najważniejsza przewaga takiej architektury nie polegałaby na wszechwiedzy; przeciwnie, jej dojrzałość przejawiałaby się w zdolności do reprezentowania własnej niepewności. Im więcej wie system, tym ważniejsze staje się pytanie, czego wiedzieć nie powinien.

Rozwój impact analytics prowadzi ostatecznie do paradoksu charakterystycznego dla całej cywilizacji danych. Z technologicznego punktu widzenia organizacje zyskują możliwość coraz dokładniejszego modelowania człowieka, łącząc dane o jego sytuacji mieszkaniowej, zdrowotnej, edukacyjnej, zawodowej, przestrzennej i behawioralnej. Taki model pozwala wcześniej rozpoznawać potrzeby i lepiej dopasowywać wsparcie, ale ta sama infrastruktura może zostać wykorzystana do nadzoru, dyskryminacji, manipulacji lub paternalistycznego podejmowania decyzji w imieniu człowieka. Nie istnieje techniczna granica, która automatycznie oddzieli troskę od kontroli. Granica ta jest aksjologiczna i instytucjonalna.

Idea Stevenson-Dodda, według której technologia powinna uwalniać ludzi od biurokracji i oddawać czas relacjom human-to-human, okazuje się ważniejsza niż jego konkretne prognozy technologiczne. Jeśli AI uwalnia pracownika od przepisywania danych, dając mu więcej czasu na rozmowę, wzmacnia funkcję społeczną instytucji. Jeśli jednak uwalnia instytucję od konieczności rozmowy z człowiekiem, bo „model już wie", prowadzi w przeciwnym kierunku. Przyszłość pomiaru wpływu nie zostanie zatem rozstrzygnięta przez pytanie, **ile informacji potrafimy uzyskać**, lecz przez to, **jaką wiedzę uznamy za legitymowaną, do jakich decyzji pozwolimy jej służyć i jakie sfery ludzkiej autonomii pozostawimy poza optymalizacją**.

To prowadzi do najgłębszego poziomu problemu. Skuteczność, additionality, SROI, Live Data, causal inference i AI odpowiadają na pytania o to, **czy i jak potrafimy wywołać zmianę**. Nie odpowiadają jednak na pytanie, **czy zmiana jest dobra, dla kogo jest dobra, kto ma prawo ją definiować i czego nawet bardzo skuteczna organizacja nie powinna robić**. Kolejna część analizy musi więc przejść od epistemologii wpływu do jego aksjologii: od mierzenia skuteczności ku godności, autonomii, sprawiedliwości, prawom człowieka i granicom społecznej optymalizacji.

> Ostatecznie największym ryzykiem cyfrowej transformacji nie jest brak danych, lecz złudzenie, że model predykcyjny może zastąpić rozmowę z człowiekiem. Jeśli AI uwolni nas od biurokracji, byśmy mogli poświęcić więcej czasu na relacje, wzmocni funkcję społeczną instytucji; jeśli jednak uznamy, że „model już wie", zamienimy troskę w nadzór. Pytanie o przyszłość pomiaru wpływu nie brzmi zatem: ile informacji potrafimy uzyskać, lecz: jakie sfery ludzkiej autonomii odważymy się pozostawić poza zasięgiem optymalizacji?

Inspiracje

[1] 

"The Difference How to Understand Measure and Maximise Social Impact" Matt Stevenson-Dodd

Practical Inspiration Publishing | ISBN: 9781788609562

Matt Stevenson-Dodd to brytyjski przedsiębiorca społeczny i specjalista w dziedzinie strategii non-profit i pomiaru wpływu społecznego. Jest założycielem i dyrektorem generalnym Trust Impact, firmy konsultingowej założonej w 2019 roku, która pomaga organizacjom charytatywnym, instytucjom sektora publicznego i przedsiębiorstwom zrozumieć, zmierzyć i zmaksymalizować ich wpływ społeczny. Przed założeniem Trust Impact, Stevenson-Dodd spędził ponad dwie dekady, kierując organizacjami non-profit, w tym pełniąc funkcję dyrektora generalnego Street League w latach 2010-2018, gdzie zrewolucjonizował działalność charytatywną w sektorze sportu dla pracowników, wprowadzając innowacyjne i przejrzyste raportowanie wpływu. Jego kluczowe osiągnięcia koncentrują się na uproszczeniu oceny wpływu społecznego, przejściu organizacji od pomiaru samej aktywności do monitorowania znaczących rezultatów życiowych oraz promowaniu transparentności w zakresie niepowodzeń organizacji. W 2020 roku został przyjęty w poczet członków Zakonu Świętego Jana za zasługi dla sektora wolontariatu.

Matt Stevenson-Dodd *is a British social entrepreneur and specialist in non-profit strategy and social impact measurement. He is the founder and chief executive officer of Trust Impact, a consultancy established in 2019 to help charities, public sector bodies, and businesses understand, measure, and maximize their social impact. Prior to founding Trust Impact, Stevenson-Dodd spent over two decades leading non-profit organizations, notably serving as chief executive officer of Street League from 2010 to 2018, where he transformed the sports-for-employment charity through innovative, transparent impact reporting. His key contributions focus on simplifying social impact evaluation, shifting organizations from measuring mere activity to tracking meaningful life outcomes, and promoting transparency regarding organizational failures. In 2020, he was admitted as a Member of the Order of Saint John for his contributions to the voluntary sector.*

Trust Impact

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:



[1] "Social Impact Investing"

Stewart Jones, Helena de Anstiss, Carmen Garcia

2021 | Routledge | ISBN: 9781000530230



[2] "Why Environmental, Social, and Governance (ESG) Is Reshaping Corporate Strategy"

Bhupendra Thapa

2026 | GRIN Verlag | ISBN: 9783389210185



[3] "New requirements on ESG reporting in Europe"

Malte Kanngießer

2024 | GRIN Verlag | ISBN: 9783389074558



[4] "Environmental, Social, and Governance (ESG) Investment and Reporting"

Michaela Bednárová, Kullapa Soratana

2025 | Springer Nature | ISBN: 9783031842351



[5] "Implementing Environmental, Social and Governance (ESG) Principles for Sustainable Businesses"

Tracy Dathe, Marc Helmold, René Dathe, Isabel Dathe

2024 | Springer Nature | ISBN: 9783031527340



[6] "Environmental, Social and Governance (ESG) Laws, Regulations and Practices in the Digital Era"

Peter Yeoh

2022 | Kluwer Law International B.V. | ISBN: 9789403541808

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 34931386-9176-49bf-acdb-da25e845ee69