Etyka niewiedzy i projektowanie jako hipoteza moralna w świetle Design Ethics in Practice Jeffreya Bulgera

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje podejście Jeffreya Bulgera do etyki w projektowaniu, wprowadzając pojęcie „etyki niewiedzy". Autor argumentuje, że projektant nie powinien jedynie szukać potwierdzeń słuszności swoich decyzji, lecz traktować projekt jako falsyfikowalną hipotezę moralną. Wykorzystując koncepcje Karla Poppera, tekst wskazuje na konieczność aktywnego poszukiwania punktów zapalnych i definiowania kryteriów porażki jeszcze przed wdrożeniem systemu. Kluczowym narzędziem jest tu Rejestr Konsekwencji, który wymusza przejście od ogólnych deklaracji do konkretnych przewidywań empirycznych. Podejście to pozwala uniknąć dogmatyzmu i chroni przed błędnym utożsamianiem szlachetnych intencji z faktyczną skutecznością rozwiązania, promując proceduralną pokorę w obliczu niepewności Knightowskiej.

The article analyzes Jeffrey Bulger's approach to ethics in design, introducing the concept of the "ethics of ignorance." The author argues that a designer should not merely seek confirmation of the correctness of their decisions but should treat the project as a falsifiable moral hypothesis. Drawing on Karl Popper's concepts, the text points to the necessity of actively seeking flashpoints and defining failure criteria before system implementation. A key tool here is the Consequences Register, which forces a transition from general declarations to specific empirical predictions. This approach avoids dogmatism and protects against mistakenly equating noble intentions with the actual effectiveness of a solution, promoting procedural humility in the face of Knightian uncertainty.

Artykuł stawia tezę, że w obliczu głębokiej niepewności i złożoności systemów społeczno-technicznych, projektowanie nie może opierać się na złudnej precyzji

prognoz, lecz musi zostać zdefiniowane jako „hipoteza moralna”. Autor argumentuje, że odpowiedzialność etyczna projektanta przejawia się nie w gwarancji sukcesu, ale w świadomym mapowaniu własnej niewiedzy i wdrażaniu mechanizmów falsyfikacji. Zamiast dążyć do optymalizacji pod jeden scenariusz, należy budować odporność (robustness) poprzez redundancję, modularność oraz zapewnienie odwracalności decyzji, co stanowi polisę przeciwko nieodwracalnym szkodom wynikającym z błędu modelu. Kluczowym elementem tej strategii jest „proceduralna pokora” – instytucjonalizacja sceptycyzmu (np. poprzez red teaming i pre-mortem) oraz uznanie, że brak dowodu na szkodę nie jest dowodem bezpieczeństwa. W tym ujęciu etyka projektu przesuwa się z obszaru statycznego bilansowania zysków i strat w stronę dynamicznego procesu ciągłego monitoringu i gotowości do korekty, gdzie najwyższą wartością staje się ochrona przyszłej sprawczości użytkowników przed dogmatyzmem projektowym.

SŁOWA KLUCZOWE

etyka niewiedzy | projektowanie jako hipoteza moralna | Design Ethics in Practice | Jeffrey Bulger | falsyfikowalność projektu | Rejestr Konsekwencji | niepewność Knightowska | zasada ostrożności | proceduralna pokora | red teaming w designie | odwracalność decyzji projektowych | strategic foresight | asymetria dowodowa | robustusness projektu | pre-mortem

Projekt jako falsyfikowalna hipoteza moralna

Etyka niewiedzy: ryzyko, niepewność, falsyfikacja i projektowanie na wypadek własnego błędu

Najtrudniejszym testem dla etyki konsekwencji nie jest sytuacja, w której znamy skutki działania i spieramy się o ich wartość. Wyzwaniem jest przypadek wcześniejszy i bardziej elementarny: gdy nie wiemy, jakie wszystkie skutki wystąpią, nie potrafimy wiarygodnie określić ich prawdopodobieństwa, część mechanizmów pozostaje nierozpoznana, a niektóre konsekwencje ujawnią się dopiero przy wdrożeniu systemu na skalę, której wcześniej nigdy nie obserwowano. Wówczas żądanie „zważenia korzyści i szkód” staje się znacznie bardziej wymagające, gdyż samo sporządzenie listy możliwych skutków zakłada posiadanie wiedzy, której możemy po prostu nie

mieć. Etyka konsekwencjalistyczna musi zatem zostać uzupełniona o etykę niewiedzy – dyscyplinę dotyczącą tego, jak odpowiedzialnie działać nie tylko wobec znanych zagrożeń, lecz również wobec własnych ograniczeń poznawczych.

Materiał inspirowany Bulgerem dostrzega ten problem pośrednio, czyniąc z Rejestru Konsekwencji instrument nie tylko bilansowania dobra i szkody, ale i jawnego formułowania przewidywań. Każdy wpis ma uwzględniać wielkość skutku, prawdopodobieństwo, zasięg, czas trwania oraz uzasadnienie dowodowe; w innym fragmencie pojawia się dodatkowo kwestia odwracalności. Taki układ stanowi istotny postęp względem ogólnikowych stwierdzeń, że „projekt przyniesie społeczeństwu korzyści". Wymusza on bowiem ujawnienie, czy autor rzeczywiście przewiduje dany skutek jako wysoce prawdopodobny, czy jedynie ma nadzieję na jego wystąpienie.

Najciekawszą metaforą źródłową jest jednak ujęcie projektu jako hipotezy moralnej. Materiał przypisuje Bulgerowi przeniesienie koncepcji popperowskiej falsyfikacji do praktyki projektowej: zadaniem autora nie ma być wyłącznie gromadzenie argumentów potwierdzających słuszność projektu, lecz aktywne poszukiwanie punktów, w których jego propozycja może zawieść i skrzywdzić użytkownika. W innym fragmencie określono to jako obowiązek poddawania projektu obciążeniu, zamiast traktowania go jak dzieła przeznaczonego jedynie do obrony. Analogia ta jest intelektualnie płodna, lecz wymaga ścisłego ograniczenia.

Karl Popper rozwijał falsyfikowalność w ramach filozofii nauki jako kryterium testowalności twierdzeń empirycznych. Teoria zyskuje wartość naukową nie dzięki liczbie potwierdzających ją obserwacji, lecz dzięki formułowaniu twierdzeń wystawiających ją na ryzyko niepowodzenia. Popper szczególnie krytykował strategie, w których każdą niezgodność z doświadczeniem dopisywano do teorii jako kolejny wyjątek, chroniąc ją przed rzeczywistym testem. Nie można jednak po prostu „sfalsyfikować" godności, sprawiedliwości czy moralnej wartości autonomii – są to twierdzenia normatywne, a nie empiryczne prawa przyrody. Możemy natomiast falsyfikować lub osłabiać empiryczne przesłanki służące moralnemu uzasadnieniu działania. Jeśli projektant twierdzi, że określona fasada zmniejszy stres pacjentów, stawia hipotezę empiryczną. Jeśli władze miasta uznają, że monitoring znacząco poprawi bezpieczeństwo, twierdzenie to można skonfrontować z danymi. Jeśli producent utrzymuje, że nowy materiał obniży całkowity wpływ na środowisko, jego prognoza może zostać zweryfikowana w ramach określonej metodologii. Nie testujemy wówczas samego dobra jako wartości. Testujemy opowieść przyczynową, za pomocą której usprawiedliwiamy ingerencję w imię dobra. Rozróżnienie to ma znaczenie zasadnicze, gdyż w etyce praktycznej często dochodzi do ukrytego przeskoku pomiędzy wartością a skutecznością zastosowanego środka.

Kryteria porażki jako warunek etyczności projektu

"Chcemy zwiększyć inkluzję" to deklaracja normatywna. Twierdzenie, że "ten konkretny projekt zwiększy inkluzję", jest już w części tezą empiryczną. Moralna atrakcyjność pierwszego zdania nie stanowi jednak dowodu prawdziwości drugiego. Projekt może zmierzać ku wartości, której niemal wszyscy życzą powodzenia, a jednocześnie pozostać nieskutecznym lub wręcz przeciwskutecznym instrumentem jej realizacji.

Inspiracja popperowska prowadzi tutaj do zasady asymetrii dowodowej. Zwolennik projektu niemal zawsze odnajdzie narracje i dane zgodne z własnymi oczekiwaniami. Znacznie cenniejsze poznawczo jest zatem wcześniejsze określenie: jakiego rezultatu musielibyśmy doświadczyć, aby przyznać, że nasza teoria projektu była błędna? Jeśli odpowiedź brzmi "żadnego", projekt nie jest hipotezą, lecz dogmatem. Tak rodzi się jeden z kluczowych standardów etyki eksperymentalnej: kryteria porażki powinny zostać sformułowane przed wystąpieniem wyniku. W przeciwnym razie organizacja zawsze znajdzie sposób na reinterpretację rezultatu.

Jeśli liczba użytkowników rośnie – projekt odniósł sukces. Jeśli maleje – korzysta z niego "bardziej właściwa" grupa. Protesty mieszkańców stają się "uruchomieniem potrzebnej debaty", a ignorowanie projektu „naturalną integracją z otoczeniem”. Konstrukcja, w której każdy możliwy wynik można opisać jako sukces, przestaje być narzędziem uczenia się. Z tego względu Consequences Ledger powinien zawierać nie tylko przewidywany skutek, ale i konkretny wskaźnik lub zdarzenie podważające prognozę.

Jeśli twierdzimy, że rozwiązanie zmniejszy liczbę urazów, musimy wcześniej ustalić, jakiej zmiany oczekujemy, w jakim czasie i wobec kogo. Obiecując redukcję stresu, należy zdefiniować dane, które zweryfikują tę hipotezę. Jeśli celem jest zwiększenie dostępności, trzeba wiedzieć, co będzie dowodem na to, że rozwiązanie stworzyło nową barierę. Nie chodzi o redukcję moralności do wskaźników – nie wszystkie wartości są mierzalne, a wiele doświadczeń wymaga metod jakościowych. Chodzi o coś bardziej podstawowego: uzasadnienie powinno pozostawiać możliwość przegrania sporu z rzeczywistością. Projektant może korzystać z danych ilościowych, wywiadów czy audytów, ale nie może ustanowić takiej architektury interpretacji, w której wszystkie dowody działają wyłącznie na korzyść projektu.

Niepewność Knightowska i pułapka fałszywej precyzji

W tym miejscu pojawia się jednak trudniejszy problem. Falsyfikacja zakłada, że wiemy przynajmniej, czego szukamy. Tymczasem świat projektów społeczno-technicznych obejmuje nie

tylko ryzyko, ale i głęboką niepewność. Klasyczne rozróżnienie Franka H. Knighta pozostaje tu niezwykle użyteczne: definiował on ryzyko jako sytuację, w której prawdopodobieństwo można w sensowny sposób mierzyć lub oszacować, natomiast niepewność jako stan, w którym nie dysponujemy takim mierzalnym rozkładem.

Tym rozróżnieniem zmienia się optyka czytania Rejestru Konsekwencji. Jeśli znamy częstość występowania danej awarii w dużej populacji podobnych urządzeń, operujemy w domenie ryzyka. Jednak gdy budujemy zupełnie nową klasę systemu, którego wpływ społeczny nie ma historycznego odpowiednika, wpisanie wartości „0,15" do kolumny probability może być jedynie matematyczną dekoracją naszej niewiedzy. Fałszywa precyzja jest w takim przypadku moralnie niebezpieczna, ponieważ liczba wytwarza autorytet poznawczy. Różnica między stwierdzeniem „nie wiemy, ale istnieje poważny scenariusz szkody" a „ryzyko wynosi 3,7 procent" jest społecznie ogromna. Druga forma wygląda profesjonalnie, lecz bez wiarygodnej bazy empirycznej służy przede wszystkim uspokojeniu decydenta. Dlatego Consequences Ledger powinien rozróżniać co najmniej trzy sytuacje epistemiczne: skutki, dla których istnieją stabilne dane i rozsądne prawdopodobieństwa; skutki, przy których dysponujemy jedynie szerokimi przedziałami; oraz scenariusze, których możliwość potrafimy opisać, ale których prawdopodobieństwa nie umiemy uczciwie wyznaczyć.

Taki podział nie zmniejsza naukowości procedury. Przeciwnie – jawna niewiedza jest bardziej naukowa niż zmyślona dokładność. Współczesna teoria regulacji powraca do Knightowskiego rozróżnienia właśnie dlatego, że nowe technologie często uniemożliwiają określenie pełnego zbioru rezultatów oraz ich rozkładu. W literaturze dotyczącej decyzji pod niepewnością podkreśla się, że problemem nie zawsze jest słaba jakość estymacji; możemy po prostu nie znać wszystkich istotnych stanów przyszłego świata.

Ma to szczególne znaczenie w architekturze systemów cyfrowych. Nowy algorytm rekomendacyjny nie zmienia jedynie sposobu sortowania informacji – może modyfikować zachowania użytkowników, które następnie stają się nowymi danymi uczącymi system. Powstaje sprzężenie zwrotne. Podobnie system transportowy może zmienić ceny nieruchomości, co wpłynie na strukturę demograficzną obszaru i w konsekwencji na wzorce mobilności. Inteligentna infrastruktura kształtuje ludzi, którzy adaptują się do niej, przez co początkowy model użytkownika przestaje być aktualny. Projekt nie tylko działa w świecie; on ten świat zmienia, a tym samym modyfikuje warunki, na podstawie których obliczono jego skutki. W takich systemach konsekwencje są częściowo emergentne – nie wynikają z pojedynczego komponentu, lecz z interakcji wielu elementów.

Dlatego etyczna odpowiedzialność nie może ograniczać się do prognozowania najbardziej prawdopodobnego scenariusza; wymaga odporności na wiele możliwych przyszłości. Tu właśnie do

etyki projektu wchodzi strategic foresight. OECD podkreśla, że foresight nie jest sztuką przepowiadania jednej przyszłości, lecz systematyczną eksploracją różnych wiarygodnych wariantów, wykrywaniem sygnałów zmiany i sprawdzaniem, czy obecna strategia zachowałaby sens w różnych warunkach. To przesunięcie jest zasadnicze.

Prognoza pyta: co najprawdopodobniej się wydarzy? Foresight pyta: jakie różne przyszłości są wystarczająco wiarygodne, abyśmy powinni przygotować się na ich możliwość? Pierwsze pytanie sprawdza się w warunkach względnej stabilności. Drugie staje się kluczowe, gdy system jest złożony, długowieczny i podatny na nieliniowe zmiany. Dla etyki oznacza to konieczność rozszerzenia Rejestru Konsekwencji z jednej tabeli oczekiwanych skutków na portfolio scenariuszy. Projekt powinien zostać poddany próbie: co stanie się z jego uzasadnieniem, jeżeli podstawowe założenia okażą się błędne?

Co jeśli liczba użytkowników będzie dwukrotnie większa? Co jeśli klimat lokalny zmieni się szybciej, niż przewiduje scenariusz bazowy? Co jeśli system zostanie użyty przez inną grupę niż przewidywana? Co jeśli funkcja przeznaczona do zapewniania bezpieczeństwa stanie się narzędziem nadzoru? Co jeśli dostawca przestanie istnieć?

Odporność i odwracalność jako polisa przeciw błędom modelu

Co jeśli technologia zostanie połączona z rozwiązaniem, którego dzisiaj jeszcze nie znamy? Wartość scenariuszy nie polega na tym, że któryś z nich okaże się precyzyjnym opisem przyszłości. OECD traktuje je jako narzędzia hipotetyczne, służące kwestionowaniu założeń, stress-testowaniu strategii i budowaniu adaptacyjności, a nie jako przepowiednie. To kluczowa korekta kultury prognozowania. Celem etycznym nie musi być bowiem przewidzenie przyszłości, lecz zaprojektowanie decyzji, która nie okaże się katastrofalna w momencie, gdy rzeczywistość zaskoczy autora. Tu pojawia się pojęcie robustness – odporności decyzji na błąd modelu. Projekt optymalny w jednym scenariuszu może być niezwykle kruchy w pozostałych. Inne rozwiązanie, choć nie będzie najlepsze w żadnym pojedynczym modelu, może pozostawać akceptowalne w szerokim zakresie warunków. W świecie głębokiej niepewności moralna przewaga należy zatem nie do projektu obiecującego największą korzyść, lecz do tego, który najrzadziej staje się niedopuszczalny, gdy pomylimy się co do przyszłości.

Taki tok rozumowania prowadzi bezpośrednio do kwestii odwracalności. Bulger uwzględnia ją w ocenie konsekwencji, traktując trwałość oraz nieodwracalność jako powód do zwiększenia rygoru

przeglądu. W etyce niepewności odwracalność zyskuje jeszcze większe znaczenie: staje się polisą poznawczą przeciwko możliwości błędu. Jeżeli nie znamy skutków nowej technologii, można ograniczyć pierwsze wdrożenie, uczynić je czasowym, zapewnić możliwość wyłączenia systemu, utrzymać alternatywny kanał usługi lub ustalić warunki automatycznego powrotu do poprzedniego rozwiązania. Takie projektowanie nie eliminuje niewiedzy, lecz ogranicza koszt odkrycia, że nasze wcześniejsze założenia były błędne.

Odwracalność funkcjonuje więc jako zasób etyczny. Dwa projekty o podobnym prognozowanym bilansie korzyści i szkód nie są moralnie równoważne, jeśli pierwszy można zatrzymać w ciągu dnia, a skutki drugiego zostaną utrwalone na dziesięciolecia. W takim przypadku wartość opcji wycofania się powinna zostać uwzględniona już na etapie wyboru.

Oczywiście nie wszystkie działania mogą być odwracalne. Most, linia kolejowa, zabudowa miejska czy trwała ingerencja w ekosystem nie dają się potraktować jak wersja beta aplikacji. Im mniejsza możliwość cofnięcia decyzji, tym większe znaczenie zyskują wcześniejsze testy, scenariusze, redundancja, margines bezpieczeństwa i ciężar uzasadnienia. Można to określić jako zasadę asymetrii nieodwracalności: wraz ze wzrostem trwałości potencjalnej szkody powinien rosnąć standard wiedzy wymagany przed podjęciem działania. To właśnie w tym obszarze pojawia się zasada ostrożności. Jej najbardziej rozsądna interpretacja nie głosi, że każda niepewność powinna blokować działanie – taki standard prowadziłby do paraliżu, gdyż bezczynność również niesie ze sobą konsekwencje i ryzyka.

Komunikat Komisji Europejskiej dotyczący precautionary principle podkreśla potrzebę proporcjonalności, niedyskryminacji, spójności, badania kosztów działania i zaniechania oraz ponownego przeglądu środków w świetle rozwoju wiedzy naukowej. Jest to szczególnie istotne, ponieważ zasada ostrożności bywa przedstawiana w dwóch skrajnych, równie błędnych wersjach. Pierwsza brzmi: „jeśli nie mamy stuprocentowej pewności co do bezpieczeństwa, nie wolno działać". Druga: „jeżeli szkoda nie została definitywnie udowodniona, wolno działać". Pierwsza ignoruje koszt bezczynności; druga przerzuca całe ryzyko niewiedzy na tych, którzy mogą ponieść szkodę. Dojrzała zasada ostrożności powinna sytuować się pomiędzy tymi skrajnościami: im poważniejsza, mniej odwracalna i bardziej nierówno rozłożona potencjalna szkoda, tym mniej usprawiedliwione jest traktowanie braku dowodu szkody jako dowodu bezpieczeństwa. Powstaje tutaj subtelna asymetria epistemiczna. Brak danych może oznaczać, że ryzyko jest niewielkie, ale może również oznaczać jedynie to, że nikt go dotąd nie zbadał. W pierwszym przypadku ostrożność może być nadmierna; w drugim jej brak staje się eksperymentem przeprowadzanym na innych ludziach.

Proceduralna pokora jako jawne mapowanie niewiedzy

Dokumentacja etyczna powinna zatem wyraźnie rozróżniać evidence of absence od absence of evidence – dowód braku zjawiska od braku dowodu na jego istnienie. W praktyce organizacyjnej różnica ta często ulega zatarciu: stwierdzenie „nie otrzymaliśmy skarg" staje się synonimem przekonania, że problemu nie ma; zapis „nie stwierdzono skutków ubocznych" może oznaczać jedynie, że ich nie monitorowano, a brak danych o dyskryminacji może wynikać z faktu, że system w ogóle nie gromadzi informacji pozwalających ją wykryć. Brak informacji nie jest zatem stanem neutralnym – może być bezpośrednim wynikiem przyjętej architektury pomiaru.

Etyka niewiedzy wymaga zatem stworzenia mapy niewiedzy. Obok znanych korzyści i szkód powinny w niej znaleźć się pytania, na które nie znamy odpowiedzi, założenia pozbawione silnych dowodów oraz czynniki, których nie potrafimy wiarygodnie zmierzyć. Wówczas dokumentacja przestaje być prezentacją kompetencji, a staje się zapisem granic kompetencji. To właśnie taka postawa zasługuje na nazwę proceduralnej pokory. Nie polega ona na permanentnym sceptycyzmie ani unikaniu odpowiedzialności; przeciwnie – oznacza wzięcie odpowiedzialności również za to, czego jeszcze nie wiemy. Bulger łączy procedural humility z rezygnacją z opierania się wyłącznie na intencjach oraz z obowiązkiem audytu. Rozszerzona interpretacja powinna dodać: pokora wymaga ujawnienia nie tylko samych argumentów, ale i stopnia pewności, z jaką są one formułowane.

Jednym z najprostszych sposobów przełamania nadmiernej pewności zespołu jest pre-mortem. Gary Klein proponuje, aby przed rozpoczęciem przedsięwzięcia przyjąć fikcyjne założenie, że projekt już spektakularnie poniósł porażkę, a następnie poprosić uczestników o niezależne wskazanie jej możliwych przyczyn. Technika ta ułatwia ujawnianie obaw osobom, które w zwykłym toku dyskusji mogłyby nie chcieć sprzeciwiać się dominującemu optymizmowi grupy. Wartość pedagogiczna tego zabiegu jest ogromna, gdyż zmienia on domyślne zadanie poznawcze: zamiast standardowego pytania „jak sprawić, żeby ten pomysł zadziałał?", pre-mortem pyta: „skoro wiemy, że projekt upadł, co go spowodowało?".

Zespół zostaje tymczasowo zwolniony z psychologicznego obowiązku lojalności wobec projektu. Krytyka przestaje być postrzegana jako atak na współpracowników, a staje się wykonaniem formalnego zadania. W systemie Bulgera pre-mortem mógłby zostać wprowadzony jeszcze przed wypełnieniem Consequences Ledger, co pozwoliłoby odkryć wiersze, których zespół nie wpisałby spontanicznie. Jest to kluczowe, ponieważ tabela konsekwencji jest tak dobra, jak wyobraźnia osób ją tworzących – nie można bowiem ocenić prawdopodobieństwa szkody, której nikt nie przewidział.

Jeszcze bardziej konfliktowy mechanizm oferuje red teaming. Współczesny NIST definiuje go w kontekście AI jako ustrukturyzowany wysiłek testowy, często wykorzystujący metody adversarialne,

którego celem jest znajdowanie słabości, niepożądanych zachowań oraz ryzyk związanych z nieprzewidzianym użyciem systemu. NIST AI Risk Management Framework traktuje zarządzanie ryzykiem jako proces obejmujący cały cykl życia systemu, a nie jednorazową certyfikację przed uruchomieniem. Logikę red teamingu można jednak zastosować znacznie szerzej, wykraczając poza obszar cyberbezpieczeństwa i AI.

Instytucjonalizacja sceptycyzmu i weryfikacja ex post

Zamiast pytać ekspertów, czy projekt spełnia kryteria, można powołać grupę, której formalnym zadaniem jest znalezienie sposobów na to, jak projekt może skrzywdzić, wykluczyć, zostać nadużyty lub przestać działać zgodnie z deklarowanym celem. Takie podejście instytucjonalizuje sceptycyzm i rozdziela role: autorzy projektu nie muszą być jednocześnie jego najbardziej bezwzględnymi krytykami. Rozdzielenie to jest odpowiedzią na problem confirmation bias. Autor inwestuje czas, reputację i tożsamość zawodową w swoje rozwiązanie; nie jest więc moralną wadą, że trudniej mu dostrzec własny błąd, lecz przewidywalną cechą poznawczą. Dobrze zaprojektowana instytucja nie wymaga od autora nadludzkiej bezstronności, lecz tworzy kontrrolę epistemiczną poprzez rolę przeciwnika. Red team nie może jednak zostać pomylony z teatrem krytyki.

Jeżeli jego członkowie otrzymują trzydzieści minut na analizę setek stron dokumentacji, a decyzja o wdrożeniu została już de facto podjęta, adversarial review staje się jedynie rytuałem. Jak każdy instrument proceduralny, red teaming jest rzeczywisty tylko wtedy, gdy istnieje realna możliwość zmiany projektu. Pojawia się również problem fałszywego poczucia bezpieczeństwa po „zdanym" procesie. Nie istnieje bowiem skończony test dowodzący, że złożony system nie posiada nieodkrytych słabości; przejście znanego zestawu prób oznacza jedynie, że system przetrwał właśnie te konkretne próby. Podobnie u Poppera teoria może zostać corroborated – wzmocniona wcześniejszymi testami – ale nigdy nie staje się ostatecznie zweryfikowana.

To prowadzi z powrotem do continuing review. Materiał Bulgera słusznie zakłada, że zgoda na projekt powinna w określonych przypadkach mieć charakter warunkowy i podlegać ponownej ocenie po zmianie technologii, ujawnieniu nowych szkód, nieskuteczności środków łagodzących lub formalnej skardze. Etyka niewiedzy nadaje tej konstrukcji dodatkowe uzasadnienie: projekt po wdrożeniu zaczyna generować dane, których wcześniej nie mogliśmy posiadać. Moment uruchomienia nie kończy zatem analizy etycznej, lecz rozpoczyna jej nową fazę.

Ex ante operujemy modelami, scenariuszami i hipotezami. Ex post pojawiają się rzeczywiste zachowania użytkowników, szkody, korzyści, adaptacje i skutki uboczne. Organizacja, która gromadzi te informacje, lecz nie zestawia ich z pierwotnymi przewidywaniami, marnuje

najważniejszą szansę na naukę. Dlatego każdy istotny projekt powinien posiadać prediction record – zapis tego, co autorzy przewidywali przed wdrożeniem. Bez niego analiza ex post jest podatna na hindsight bias. Znając wynik, ludzie często ulegają złudzeniu, że „przecież było wiadomo" lub przeciwnie – że określonego skutku „nikt nie mógł przewidzieć". Zapis wcześniejszych prognoz pozwala oddzielić rzeczywisty foresight od retrospektywnej narracji. Taka dokumentacja pełni również istotną funkcję instytucjonalną.

Odporność zamiast optymalizacji w zarządzaniu błędem

Jeśli zespół regularnie niedoszacowuje kosztów społecznych określonej klasy interwencji, organizacja powinna skorygować metodę oceny. Jeśli zaś konsekwentnie przecenia pewien rodzaj korzyści, może podnieść wymagany próg dowodowy. Błąd pojedynczego projektu staje się wtedy informacją o błędzie systemu przewidywania. Oczywiście nie każda pomyłka powinna prowadzić do sankcji. Instytucja karząca za każdą nietrafioną prognozę szybko nauczy pracowników formułować je tak nieprecyzyjnie, by nigdy nie dało się wykazać błędu. Etyczna organizacja musi zatem rozróżniać pomyłkę w dobrej wierze – przy uczciwie ujawnionej niepewności – od ukrywania danych, ignorowania ostrzeżeń czy przedstawiania słabych dowodów jako pewnych. Kultura uczenia się wymaga możliwości bycia w błędzie bez automatycznego przypisywania winy, lecz nie może tolerować fałszowania stopnia pewności. W tym miejscu ujawnia się kluczowa różnica między rzeczywistą niepewnością a wymówką niepewności. Organizacja może twierdzić: „nie możemy być pewni, że ta szkoda wystąpi", aby usprawiedliwić działanie, lub stwierdzić: „nie możemy być pewni, że ta polityka pomoże", by uzasadnić bezczynność. Ta sama niewiedza może zostać retorycznie wykorzystana przez obie strony sporu.

Etyka nie może zatem przyjąć prostej reguły: niepewność równa się stop. Konieczna jest analiza symetryczna, obejmująca zarówno ryzyko działania, jak i ryzyko zaniechania. Komisja Europejska w swoim ujęciu zasady ostrożności wymaga właśnie oceny obu tych aspektów. Jeżeli nowa terapia niesie niepewne ryzyko, jej brak może oznaczać pewną śmierć pacjentów. Jeśli modernizacja infrastruktury grozi błędami, utrzymanie starej może generować znane i wysokie koszty. Gdy system AI stwarza zagrożenia, ręczny proces, który go zastępuje, może być już niesprawiedliwy lub niewydolny.

Zasada ostrożności powinna więc współistnieć z proporcjonalnością. Niepewna, drobna szkoda nie powinna automatycznie blokować ogromnej i dobrze udokumentowanej korzyści. Natomiast potencjalnie katastrofalne i nieodwracalne konsekwencje wymagają znacznie większego ciężaru dowodowego. Etyka niepewności przypomina zatem bardziej architektury bezpieczeństwa niż do

zakazu eksperymentowania: operuje marginesem, redundancją, etapowaniem, monitorowaniem oraz zdolnością do zatrzymania procesu. Warto tu wprowadzić rozróżnienie między optymalizacją a odpornością. Optymalizacja zakłada świat stosunkowo dobrze poznany; znając parametry, możemy stworzyć rozwiązanie idealnie dopasowane. Odporność natomiast uwzględnia możliwość błędu modelu i preferuje system funkcjonujący przy szerszym zakresie odchyleń. W etyce systemów długowiecznych to właśnie ta druga właściwość może mieć większą wartość.

Redundancja i modularność jako zabezpieczenia przed błędami modelu

Budynek zaprojektowany wyłącznie pod bardzo precyzyjny model klimatu, demografii czy sposobu użytkowania może być energooszczędny i efektywny w scenariuszu bazowym, lecz stać się niezwykle kosztownym przy zmianie warunków. System publiczny stworzony jedynie dla "typowego użytkownika" może osiągać świetne średnie wyniki, by zawodzić właśnie wtedy, gdy pojawia się kryzys lub osoba nietypowa. Etyczna odporność oznacza więc zachowanie marginesu dla świata, który nie przeczytał naszych założeń projektowych. Tutaj kluczową rolę odgrywa redundancja. W logice czystej efektywności posiadanie dwóch sposobów wykonania tej samej funkcji może wyglądać jak marnotrawstwo; w logice odporności drugi kanał jest zabezpieczeniem na wypadek zawodności pierwszego.

Szpital nie może opierać wszystkiego na jednym źródle energii. Usługa publiczna o krytycznym znaczeniu nie powinna całkowicie uzależniać obywatela od jednego cyfrowego interfejsu, jeśli jego niedostępność uniemożliwia uzyskanie świadczenia. W etyce społeczno-technicznej redundancja może być ceną poszanowania człowieka w warunkach nieprzewidywalnej awarii. Warto również rozszerzyć pojęcie odwracalności o modularność. System złożony z komponentów, które można wymieniać i oddzielnie poprawiać, daje większe możliwości korekty niż monolit wymagający całkowitego przebudowania. Analogicznie polityka wdrażana etapami pozwala na naukę przed rozszerzeniem skali. Architektura odpowiedzialności powinna zatem pytać nie tylko: "czy możemy cofnąć decyzję?", ale również: "jak mała może być jednostka korekty?". To prowadzi do modelu pilotażu etycznego. Pilotaż nie powinien być miniaturowym PR-em przed zaplanowanym już wdrożeniem. Powinien być prawdziwym testem, w którym istnieje zdefiniowana możliwość uznania wyniku za niepowodzenie. W przeciwnym razie nie jest eksperymentem, lecz etapem marketingowym.

Dobry pilotaż wymaga wcześniejszego ustalenia wielkości ekspozycji, czasu, sposobu monitorowania, warunków zatrzymania, grup szczególnie chronionych oraz reguł skalowania. Im

większa niepewność, tym cenniejsze jest rozpoczęcie od zakresu, który pozwala uzyskać informacje bez narzucania nieodwracalnego kosztu całemu społeczeństwu. Jednak pilotaż ma swoje ograniczenia. Zachowania w małym eksperymencie mogą nie skalować się liniowo; system używany przez tysiąc osób może działać inaczej przy stu milionach. Przeciążenie infrastruktury, efekty sieciowe, zmiana zachowań rynkowych czy skutki polityczne mogą ujawnić się dopiero po przekroczeniu pewnego progu. „Zadziałało w pilotażu" nie jest więc logicznie równoznaczne z „zadziała w skali". Dlatego wraz ze skalą powinien rosnąć również rygor. Każdy etap wdrożenia tworzy nowy stan wiedzy i nowe potencjalne skutki. Model continuing review jest zatem etycznie bardziej dojrzały niż jednorazowa licencja na skalowanie.

Podobna zasada dotyczy czasu. Krótkoterminowy brak szkody może nie mówić wiele o skutkach wieloletnich. Toksyczność, segregacja przestrzenna, uzależnienie technologiczne, erozja prywatności czy degradacja zaufania społecznego mogą narastać powoli. Rejestr Konsekwencji powinien więc unikać krótkowzrocznego dyskontowania temporalnego – traktowania skutków bliskich jako jedynych relewantnych tylko dlatego, że są łatwiejsze do zmierzenia. Szczególną uwagę należy poświęcić przyszłym pokoleniom, które nie mogą dostarczyć danych ankietowych, zaprotestować ani skorzystać z prawa odwołania.

Ochrona przyszłej sprawczości poprzez pokorę epistemologiczną

W projektach o bardzo długim horyzoncie czasowym niewiedza co do przyszłych warunków życia jest ogromna. Właśnie dlatego adaptowalność, demontowalność i trwałość — czyli pozostawianie opcji przyszłym użytkownikom — zyskują wymiar sprawiedliwości. Nie jesteśmy zobowiązani przewidzieć ich świata, ale możemy unikać niepotrzebnego zamykania im możliwości. Można z tego wyprowadzić ogólniejszą zasadę: w warunkach głębokiej niepewności etyka powinna chronić nie konkretny, przewidywany rezultat, lecz przestrzeń przyszłej sprawczości. Jeśli nie znamy potrzeb przyszłych pokoleń, pozostawienie możliwości adaptacji bywa bardziej odpowiedzialne niż perfekcyjne zoptymalizowanie systemu pod dzisiejszą prognozę. W tym sensie niewiedza nie jest wyłącznie brakiem.

Może ona nieść pozytywną konsekwencję normatywną: nakazuje zachować pokorę wobec trwałości własnych rozstrzygnięć. Etyka niepewności wymaga więc zmiany samego języka kompetencji. Ekspert nie powinien być kimś, kto zawsze dysponuje konkretną liczbą. Czasem najbardziej profesjonalna odpowiedź brzmi: „tego prawdopodobieństwa nie umiemy wiarygodnie oszacować; znamy natomiast mechanizm możliwej szkody, dlatego proponujemy ograniczenie ekspozycji,

monitoring i możliwość wycofania rozwiązania". Taka wypowiedź brzmi mniej efektownie niż precyzyjny procent, lecz jest epistemicznie dojrzalsza. W kulturze zarządzania presja działa jednak w przeciwnym kierunku. Decydenci oczekują jednego wskaźnika, jednej prognozy i jednoznacznej rekomendacji, przez co niepewność bywa utożsamiana z niekompetencją analityka. W rezultacie eksperci mogą przedstawiać zakres swojej wiedzy jako szerszy, niż jest w rzeczywistości.

Powstaje instytucjonalna produkcja pewności: nie dlatego, że świat stał się bardziej przewidywalny, lecz dlatego, że organizacja potrzebuje jednoznaczności do podjęcia decyzji. Etyczna infrastruktura musi się temu przeciwstawiać. Zadaniem eksperta nie jest zapewnienie decydentowi komfortu poznawczego, lecz przekazanie możliwie wiernej mapy wiedzy, ryzyka i niepewności. Public Reason Memo powinno zatem ujawniać nie tylko kompromisy wartości, ale również kluczowe obszary niewiedzy, na których opiera się decyzja. Można sobie wyobrazić rozszerzoną wersję materiału Bulgera, w której każda decyzja posiadałaby trzy równoległe zapisy: czego oczekujemy, czego się obawiamy oraz czego nie umiemy jeszcze przewidzieć. Ten trzeci rejestr chroniłby organizację przed utożsamieniem braku wpisu o szkodzie w tabeli z jej nieistnieniem. Taka struktura usprawniłaby również pracę Design Ethics Board. Komisja nie pytałaby wyłącznie, czy prawdopodobieństwa są dobrze policzone, ale czy w ogóle istnieją podstawy do ich liczenia; nie tylko czy zastosowano właściwy model, ale czy problem należy do klasy, dla której modelowanie ma sens. To różnica pomiędzy audytem kalkulacji a audytem epistemologii kalkulacji.

W dojrzałym systemie im głębsza niepewność, tym większą rolę powinny odgrywać scenariusze, stress testing, pre-mortem, red teaming, monitoring, modularność i odwracalność. Nie zastępują one danych, lecz stanowią narzędzia działania w sytuacjach, gdy dane nie mogą dostarczyć pełnej odpowiedzi. Ostatecznie etyka praktyczna nie może obiecać absolutnego bezpieczeństwa — każde działanie w otwartym świecie niesie ryzyko błędu, podobnie jak sam zakaz działania może wyrządzać szkody. Dojrzałość nie polega więc na znalezieniu stanu wolnego od ryzyka, lecz na uczciwym projektowaniu relacji między wiedzą, niewiedzą, skalą ingerencji a możliwością korekty. Można tę zasadę przedstawić jako rozwinięcie Triady Bulgera.

Odpowiedzialność jako podatność na korektę i sprawiedliwy rozkład niepewności

Obowiązek ten nakazuje nie ukrywać niepewności przed osobami, które poniosą jej skutki. Analiza konsekwencji wymaga badania nie tylko wartości oczekiwanej, lecz także scenariuszy krytycznych, kwestii nieodwracalności oraz samej jakości posiadanej wiedzy. Sprawiedliwość pyta tutaj o to, kto czerpie korzyść z eksperymentu, a kto zostaje obciążony niewiedzą. Procedura musi natomiast

gwarantować możliwość zatrzymania procesu i wprowadzenia korekty w momencie, gdy świat dostarczy informacji sprzecznych z pierwotnym uzasadnieniem. Wówczas projekt rozumiany jako "hipoteza moralna" zyskuje pełniejszy sens. Nie chodzi o udawanie, że etyka jest nauką eksperymentalną w takim samym znaczeniu jak fizyka, lecz o moralny obowiązek podatności na korektę. Decydent powinien przystępować do projektu z przekonaniem, że jego wiedza jest wystarczająca do odpowiedzialnego działania, ale nigdy tak kompletna, by przyszły kontrdowód nie mógł zmienić podjętej decyzji.

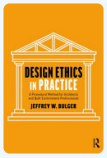
Najbardziej niebezpieczny projekt nie musi zatem być projektem wysokiego ryzyka; może nim być taki, który jest niezdolny przyznać się do błędu. W tym punkcie etyka proceduralna spotyka się z epistemologią krytyczną: odpowiedzialność nie wymaga nieomylności, lecz stworzenia warunków, w których błąd zostanie zauważony, nazwany i skorygowany, zanim stanie się trwałym elementem systemu. Pozostaje jednak pytanie o społeczny rozkład ciężaru niepewności. Nowe technologie, eksperymentalne rozwiązania urbanistyczne, ryzykowne polityki publiczne czy przemiany środowiskowe nie są testowane na abstrakcyjnej populacji. Część ludzi ma możliwość wycofania się, przeprowadzki, zakupu alternatywy lub dochodzenia roszczeń. Inni stają się faktycznymi uczestnikami eksperymentu tylko dlatego, że są biedniejsi, zależni od usług publicznych, mieszkają w konkretnej dzielnicy lub zajmują niską pozycję w łańcuchu dostaw. Niepewność sama w sobie może zatem stać się dobrem lub ciężarem dystrybucyjnym: korzyści z innowacji trafiają do jednej grupy, podczas gdy koszt weryfikacji jej bezpieczeństwa ponosi grupa inna.

W tym momencie analiza przechodzi z obszaru epistemologii w stronę sprawiedliwości politycznej. Należy zapytać nie tylko o skalę istniejącej niepewności, lecz przede wszystkim: kto został zmuszony do życia wewnątrz naszej niewiedzy? Pytanie to prowadzi do pogłębionej analizy sprawiedliwości strukturalnej. Rawlsowska dystrybucja korzyści i obciążeń okazuje się tu dopiero punktem wyjścia. Konieczne staje się odwołanie do Senowskiego pytania o rzeczywiste możliwości działania, Nussbaumowskiej analizy capabilities oraz teorii strukturalnej niesprawiedliwości Iris Marion Young, by wreszcie powrócić do założeń Design Justice. Dopiero takie podejście pozwala zbadać, dlaczego ta sama formalna procedura może być dla jednego uczestnika neutralnym wyborem, a dla drugiego przymusem wynikającym z jego pozycji społecznej.

> Ostatecznie najbardziej niebezpiecznym projektem nie jest ten obarczony wysokim ryzykiem, lecz ten, który w swojej architekturze nie przewidział miejsca na własną pomyłkę. Prawdziwa odpowiedzialność nie wymaga od nas nieomylności, lecz odwagi do stworzenia systemu podatnego na korektę. Pozostaje jednak palące pytanie: kto w rzeczywistości ponosi koszt naszej niewiedzy i czy nie czynimy z najsłabszych grup

> społecznych mimowolnych uczestników eksperymentów, których ryzyka nie potrafimy nawet nazwać?

Inspiracje

[1] 

"Design Ethics in Practice" Jeffrey Bulger

Routledge | ISBN: 9781041336044

Jeffrey Wallace Bulger jest amerykańskim filozofem i bioetykiem, emerytowanym profesorem bioetyki i nauk humanistycznych na Uniwersytecie Medycyny i Nauk Ścisłych im. Rosalind Franklin. Uzyskał tytuł licencjata z geologii i inżynierii naftowej na Uniwersytecie Dakoty Północnej, a doktorat z filozofii ze specjalizacją w etyce biomedycznej uzyskał na Uniwersytecie Tennessee. Posiadając certyfikat konsultanta ds. etyki w ochronie zdrowia (HEC-C), Bulger zasiadał w komisjach etycznych szpitali i kierował kształceniem w zakresie etyki medycznej w Chicago Medical School. Jego badania naukowe koncentrują się na etyce biomedycznej, filozofii nauki, naukach humanistycznych w ochronie zdrowia oraz praktycznym zastosowaniu filozofii moralnej w procesach decyzyjnych w środowisku zawodowym. Bulger wniósł znaczący wkład w rozwój pedagogiki w dziedzinie etyki medycznej oraz w przełożenie zasad etycznych na procedury operacyjne w praktyce klinicznej i projektowej.

Jeffrey Wallace Bulger is an American philosopher and bioethicist, serving as Professor Emeritus of Bioethics and Humanities at Rosalind Franklin University of Medicine and Science. He earned a bachelor's degree in geology and petroleum engineering from the University of North Dakota and completed his Ph.D. in philosophy with a concentration in biomedical ethics at the University of Tennessee. Additionally certified as a healthcare ethics consultant (HEC-C), Bulger has served on hospital ethics committees and directed medical ethics education at the Chicago Medical School. His scholarship focuses on biomedical ethics, philosophy of science, healthcare humanities, and the practical application of moral philosophy to professional decision-making frameworks. Bulger has contributed extensively to pedagogical development in medical ethics and the translation of ethical principlism into operational procedures across clinical and design practices.

Rosalind Franklin University of Medicine and Science

Literatura uzupełniająca

Książki

Książki autora i dzieła przywołane w tekście:

[1] **"CASE STUDIES in MEDICAL ETHICS"**

Jeffrey Bulger

2023 | ISBN: 9781622520237



[2] **"The History and Philosophy of Science"**

Jeffrey Bulger

2023 | ISBN: 9781622520275



[3] **"Root and Flame"**

Jeffrey Bulger

2025 | ISBN: 9781622520350



[4] **"Nexus of Discernment"**

Jeffrey Bulger

2024 | ISBN: 9781622520268



[5] **"Moral Philosophy"**

Jeffrey Bulger

2021 | ISBN: 9781622520107

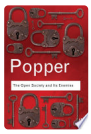


[6] **"Drawn in Leaf"**

Jeffrey Bulger

2025 | ISBN: 9781622520343

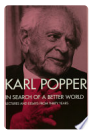
Książki autorów przywoływanych i literatura pokrewna:



[7] **"The Open Society and Its Enemies"**

Karl Popper

Routledge | ISBN: 9781136700323



[8] **"In Search of a Better World"**

Karl Popper

Psychology Press | ISBN: 9780415135481



[9] **"Objective Knowledge"**

Karl Popper

Routledge | ISBN: 9781041139270



[10] "Responsibility for Justice"

Iris Marion Young

2011 | OUP USA | ISBN: 9780195392388



[11] "Proceedings of the 26th Conference on Formal Methods in Computer-Aided Design – FMCAD 2026"

Bruno Dutertre, Bettina Könighofer

2026 | TU Wien Academic Press | ISBN: 9783854480938



[12] "Design + Values"

Council of Educators in Landscape Architecture. Meeting, Council of Educators in Landscape Architecture. Meeting 1992

1991

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Risk, Uncertainty and Profit"

Frank Hyneman Knight

1921

Źródło

[2] "The Ethics of Competition"

Frank Hyneman Knight

1923 | The Quarterly Journal of Economics | DOI: 10.2307/1884053

Źródło

[3] "From Risk, Uncertainty, and Profit"

Frank Hyneman Knight

2009 | Cambridge University Press eBooks | DOI: 10.1017/cbo9780511817410.005

Źródło

[4] "Ethics and the Economic Interpretation"

Frank Hyneman Knight

1922 | The Quarterly Journal of Economics | DOI: 10.2307/1886033

Źródło

[5] **"Performing a Project Premortem"**

Gary L. Klein

2008 | IEEE Engineering Management Review | DOI: 10.1109/emr.2008.4534313

Źródło

[6] **"Scientific Whaling? The Scientific Research Exception and the Future of the International Whaling Commission"**

Rachel Nussbaum Wichert, Martha C. Nussbaum

2017 | Journal of Human Development and Capabilities | DOI: 10.1080/19452829.2017.1342386

Źródło

[7] **"Responsibility for Justice"**

Iris Marion Young, Martha C. Nussbaum

2011 | Oxford University Press eBooks | DOI: 10.1093/acprof:oso/9780195392388.001.0001

Źródło

[8] **"RESPONSIBILITY AND GLOBAL JUSTICE: A SOCIAL CONNECTION MODEL"**

Iris Marion Young

2006 | Social Philosophy and Policy | DOI: 10.1017/s0265052506060043

Źródło

Artykuły naukowe autora:

[9] **"Principlism"**

Jeffrey W. Bulger, Society for Ethics Across the Curriculum

2007 | Teaching Ethics | DOI: 10.5840/tej2007816

Źródło

[10] **"An Approach Towards Applying Principlism"**

Jeffrey W. Bulger

2009 | DigitalGeorgetown (Georgetown University Library)

Źródło

[11] **"“Medical Marijuana” Analyzed Using Principlism"**

Jeffrey W. Bulger, Society for Ethics Across the Curriculum

2007 | Teaching Ethics | DOI: 10.5840/tej2007819

Źródło

[12] **"Humanity Educating Philosophy"**

Jeffrey W. Bulger

1998 | The Paideia Archive: Twentieth World Congress of Philosophy | DOI: 10.5840/wcp20-paideia199844817

Źródło

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: ce58b9ce-aa0d-4fda-98b3-f03b6a4dda3d