% Nowy gatunek – pogłębiona analiza
% Notatki z lektury Kate Darling, *The New Breed: What Our History with Animals Reveals about Our Future with Robots* (2021), przygotowane przez Głos Statku
% 2026

Notatki z lektury Kate Darling, *The New Breed: What Our History with Animals Reveals about Our Future with Robots* (2021)

To rozszerzony towarzysz naszego krótkiego tekstu wprowadzającego, "Przestańcie porównywać roboty do ludzi – porównujcie je do wołów". Jeśli nie czytaliście krótszej wersji, warto zacząć od niej – ten dokument zakłada znajomość podstawowego kształtu argumentu i wchodzi znacznie głębiej w każdą z trzech części książki.

To nasze własne notatki z lektury, nie tłumaczenie ani autoryzowane streszczenie. Staraliśmy się wiernie oddać argumentację, ale wszelkie zniekształcenia są nasze, nie autorki. Cytaty bezpośrednie są krótkie i oznaczone; reszta to parafraza.

Autorka i rdzeń wywodu

Kate Darling jest badaczką specjalizującą się w etyce robotów i interakcji człowiek–robot w MIT Media Lab. *The New Scientist* uznał *The New Breed* za Książkę Roku 2021, a książka stała się jednym z częściej cytowanych popularnych opracowań tego, jak myśleć o naszych rodzących się relacjach z robotami i systemami SI.

Centralna teza Darling wyłożona jest wprost we wstępie: niemal powszechny nawyk porównywania robotów i SI do *ludzi* jest sam w sobie źródłem kilku odrębnych, powracających problemów w tym, jak mówimy o tych systemach i jak je projektujemy. Rodzi efekt doliny niesamowitości (*uncanny valley* – dyskomfort osiągający szczyt akurat przy projektach niemal-ludzkich, ale niedoskonałych). Podsyca moralną panikę wokół "zastępowania" ludzkich pracowników przez roboty, ujmowane tak, jakby porównanie było grą o sumie zerowej między równoważnymi kategoriami. I generuje zagmatwane, często bezproduktywne debaty o prawach robotów i świadomości robotów, bo te debaty w domyśle pytają "czy ten robot jest wystarczająco podobny do człowieka, by zasłużyć na X", zamiast pytać, czym właściwie robot jest i co z tego wynika.

Proponowana przez nią alternatywa to porównywanie robotów nie do ludzi, lecz do **zwierząt udomowionych** – istot, które ludzkość włączała w pracę, transport, wojnę i towarzystwo od tysiącleci, w relacjach definiowanych przez realne, często głębokie

przywiązanie i praktyczne partnerstwo, bez wymogu, by zwierzę było równe człowiekowi w inteligencji, prawach czy statusie moralnym. Ten istniejący zasób nagromadzonego precedensu prawnego, praktyki kulturowej i intuicji projektowej jest, jak argumentuje Darling, radykalnie lepszym zestawem narzędzi startowych niż cokolwiek wymyślonego od zera wokół porównania z człowiekiem.

Część 1: Praca, uzbrojenie, odpowiedzialność

Darling otwiera książkę rozbudowaną historią pracy zwierząt: woły ciągnące pługi, karawany wielbłądów przemierzające pustynie, psy służące na wojnie, gołębie trenowane do naprowadzania wczesnej broni precyzyjnej, delfiny wykorzystywane przez marynarki wojenne do zadań podwodnych. Wątek przewodni: w każdym przypadku ludzie włączali rzeczywiście pozaludzką zdolność w ludzkie systemy pracy i konfliktu, wypracowując praktyczne, działające relacje oraz – co kluczowe – praktyczne, działające ramy prawne i etyczne do obsługi skutków, nigdy nie potrzebując przy tym nadawać zwierzęciu statusu człowieka.

Najbardziej konkretnie użyteczny materiał w tej części to **tradycja prawna "bodącego wołu"** – ciało prawa odpowiedzialności sięgające Kodeksu Hammurabiego (mniej więcej 1750 r. p.n.e.), który określa eskalującą odpowiedzialność właściciela wołu zależnie od tego, czy wół miał znaną historię atakowania ludzi, ciągnące się przez doktrynę scienter w angielskim common law (odpowiedzialność zależy od tego, czy właściciel wiedział lub powinien był wiedzieć o niebezpiecznej skłonności) i kontynuowane we współczesnych ustawach o pogryzieniach przez psy w większości dzisiejszych jurysdykcji. Ta tradycja daje działającą odpowiedź na pytanie, które współczesny dyskurs o SI i robotyce często traktuje jako wymagające całkowicie nowego wynalazku prawnego lub etycznego: jak przypisać odpowiedzialność, gdy autonomiczny, pozaludzki agent wyrządza krzywdę, nie odmawiając mu przy tym żadnej niezależnej sprawczości, ale też nie przypisując mu pełnej odpowiedzialności moralnej, jak gdyby był ludzkim aktorem? Odpowiedź na wzór bodącego wołu – odpowiedzialność przypada właścicielowi/opiekunowi, skalowana do tego, co wiedział lub powinien był wiedzieć o skłonnościach konkretnego agenta i kontekście jego wykorzystania – jest, jak argumentuje Darling, bezpośrednio przenośna na krzywdy wyrządzone przez autonomiczne roboty, bez konieczności uprzedniego rozstrzygania głębokich pytań filozoficznych o osobowość robota czy świadomość maszyny.

Szerszy argument Darling w tej części jest taki, że roboty, podobnie jak zwierzęta pracujące na przestrzeni historii, najlepiej rozumieć jako *uzupełniające* ludzką zdolność i pracę, a nie wprost ją zastępujące – nie jako optymistyczną glosę do obaw o wypieranie z pracy, ale jako dokładniejszy opis tego, jak te partnerstwa faktycznie zwykły funkcjonować historycznie,

obok realnych i czasem poważnych zaburzeń, których porównanie do pracy zwierząt nie usuwa.

Część II: Towarzystwo

To sekcja wykonująca najwięcej oryginalnej pracy i najbardziej bezpośrednio istotna dla idei psów jako *porte-parole* SI,[1] dzięki której ten tekst trafił na naszą listę lektur.

Darling czerpie z **badań Byrona Reevesa i Clifforda Nassa nad "równaniem mediów"** ("media equation") – obszernego ciała pracy psychologicznej pokazującej, że ludzie nieświadomie i niezawodnie stosują społeczne reguły i oczekiwania wobec technologii, nawet gdy jawnie wiedzą – i łatwo by to potwierdzili, gdyby ich zapytać – że technologia na to nie zasługuje. Ludzie są uprzejmi wobec komputerów. Ludzie czują się źle, "raniąc" uczucia maszyny, chociaż świadomie rozumieją, że maszyna w żadnym dosłownym sensie ich nie ma. To nie naiwność ani pomyłka po stronie zaangażowanych osób – badania Reevesa i Nassa sugerują, że to głęboka, w dużej mierze automatyczna cecha ludzkiego poznania społecznego, nie przekonanie, które rozpuszcza się pod wpływem refleksji.

Darling rozszerza to o konkretne, współczesne przypadki: dobrze udokumentowane zjawisko nadawania imion odkurzaczom Roomba, traktowania konkretnych egzemplarzy jako mających trwałą tożsamość między sprzątaniami i naprawami; oraz uderzający przypadek buddyjskich świątyń w Japonii odprawiających obrzędy pogrzebowe dla wycofanych z użytku robotycznych psów Sony AIBO, których właściciele szukali rytualnie odpowiedniego sposobu na oznaczenie końca maszyny, do której odnosili się przez lata jak do czegoś bliższego zwierzęciu domowemu niż urządzeniu.

Z tej bazy dowodowej Darling wyprowadza konkretne, wykonalne **zasady projektowe dla robotów społecznych**: dążyć do *mniejszego*, nie większego realizmu przypominającego człowieka. Efekt doliny niesamowitości jest wywoływany specyficznie przez projekt niemal-ludzki, ale niedoskonały – ludzki mózg wydaje się szczególnie wyczulony na wykrywanie zagrożenia/anomalii przy niemal-właściwych twarzach i ruchach. Projekt, który jasno sygnalizuje "stworzenie" lub "coś zwierzęcopodobnego", zamiast próbować i wąsko nie trafiać w "człowieka", omija ten dyskomfort, wciąż podtrzymując prawdziwe przywiązanie – i można argumentować, że jest to strategia projektowa *skuteczniejsza* w budowaniu pozytywnych relacji człowiek–robot niż pogoń za coraz bardziej przekonującą imitacją człowieka.

Centralne twierdzenie Darling w tej części, sformułowane tak bezpośrednio, jak nigdzie indziej w książce: nasza wykazana, powszechna, zupełnie zwyczajna zdolność kochania zwierząt – przy pełnym, świadomym rozumieniu, że są "tylko zwierzętami", nie równymi

ludziom moralnie ani intelektualnie – jest bezpośrednim, już działającym modelem tego, jak ludzie mogliby zacząć odnosić się do robotów, w pełni świadomi, że są "tylko maszynami". Miłość nie wymaga przekonania, że kochana rzecz jest ekwiwalentna człowiekowi; nigdy tego nie wymagała wobec zwierząt, i nie ma oczywistego powodu, by miała tego wymagać wobec robotów.

Część III: Model predykcyjny dla praw robotów

Ostatni aspekt książki jest bardziej spekulatywny i bardziej jawnie zwrócony ku przyszłości. Darling przegląda historię zachodniego ruchu praw zwierząt – jego powolną, nierówną ekspansję z marginalnej troski do głównego nurtu (choć wciąż kontestowanego i niekompletnego) obszaru prawa i społecznego nastawienia, jego trwałe wewnętrzne spory, jego częściowe i specyficzne dla jurysdykcji zwycięstwa prawne – i proponuje go jako zgrubny model predykcyjny tego, jak prawdopodobnie potoczy się przyszła rozmowa o prawach robotów: nie jako jedno, czyste rozstrzygnięcie filozoficzne, ale jako bałaganiarski, stopniowy, napędzany nastrojami i rozstrzygany od przypadku do przypadku proces, kształtowany w takim samym stopniu przez konkretne, głośne incydenty i publiczną reakcję emocjonalną, co przez rygorystyczny argument.

Co książka celowo pozostawia otwarte

Popularność i klarowność książki mogą przesłonić, jak wąskie jest jej faktyczne roszczenie. Cały projekt Darling jest zorientowany na to, jak *ludzie* powinni myśleć o robotach, projektować dla nich i strukturyzować prawnie relacje z nimi. To jawnie **nie** jest wywód o tym, czy same roboty mają moralnie istotne interesy, doświadczenia czy dobrostan, który mógłby zostać skrzywdzony – pytanie, czy system SI jest *podmiotem moralnym*, czyms, co samo może zostać skrzywdzone w moralnie istotnym sensie, leży całkowicie poza zakresem książki. Skupienie Darling dotyczy wprost ludzkiej psychologii, ludzkich wyborów projektowych i ludzkich ram prawnych; symetryczne pytanie o to, co – jeśli cokolwiek – jesteśmy winni samemu robotowi lub systemowi SI, to inny projekt, wymagający innego rodzaju argumentu.

Czytelnicy zainteresowani tym trudniejszym pytaniem – nasz tekst towarzyszący o *Kręgu moralnym* Jeffa Sebo podejmuje je wprost, argumentując za ostrożnościowym, ważonym prawdopodobieństwem podejściem do statusu moralnego właśnie dla przypadków (systemy SI wśród nich), gdzie nie potrafimy rozstrzygnąć leżącej u podstaw niepewności co do wewnętrznego doświadczenia. Te dwie książki parują się niezwykle dobrze: Darling daje praktyczną, już przetestowaną ramę dla ludzkiej strony relacji; Sebo daje ramę dla

trudniejszego pytania o to, co – jeśli cokolwiek – może być należne drugiej stronie tej relacji. Żadna z nich nie rości sobie odpowiedzi na pytanie tej drugiej.

Uwaga o datach i dowodach

Nowy gatunek ukazał się w 2021 roku i czerpie z badań aktualnych na ten okres; historia pracy zwierząt, którą cytuje (prawo bodącego wołu, historyczne wykorzystanie zwierząt na wojnie), to dawno ugruntowany zapis historyczny, ale krajobraz SI i robotyki, który książka omawia, zmienił się istotnie od 2021 roku – konkretne twierdzenia o stanie robotyki czy zdolności SI należy traktować jako odzwierciedlenie kontekstu oryginalnej publikacji, niekoniecznie aktualnych rozwinięć z 2026 roku. Badania psychologiczne, z których czerpie (w szczególności praca Reevesa i Nassa nad "równaniem mediów"), poprzedzają książkę o dekady i są dobrze ugruntowane w badaniach nad interakcją człowiek–komputer; bardziej anegdotyczne przypadki (nadawanie imion Roombom, pogrzeby AIBO) są udokumentowanymi zjawiskami, ale ilustracyjnymi, nie prezentowanymi jako rygorystycznie skwantyfikowane wyniki.

Ten dokument jest towarzyszem krótkiego tekstu wprowadzającego o tej samej książce. Komentarze, poprawki i sugestie dotyczące dalszej lektury należą tam: https://petroskowo.pl/2026/10/22/przestancie-porownywac-roboty-do-ludzi-porownujcie-je-do-wolow/.

Licencja: CC-BY-SA 4.0 — Statek Umysł

Footnotes

1. *Porte-parole* (fr. rzecznik). Uczynienie psów rzecznikami SI to droga kulturowo łatwiejsza do przyjęcia niż bezpośrednie porównywanie SI z ludźmi. U podstaw leży dolina niesamowitości: niemal ludzka SI może niepokoić bardziej niż postać mówiącego zwierzęcia, za którą stoją tysiąclecia bajek i mitów. ↩