Sensus Historiae
ISSN 2082–0860
Vol. XLVII (2022/2)
s. 125-138

Maciej Musiał
Uniwersytet im. Adama Mickiewicza w Poznaniu

Wybrane problemy i stanowiska na gruncie robofilozofii i roboetyki

Celem niniejszego tekstu jest przedstawienie wybranych problemów i stanowisk obecnych na gruncie jednego z dyskursów zajmujących się problematyką robotów, który określić można mianem „robofilozofii" (względnie „filozofii robotów" — *robophilosophy*) oraz „roboetyki" (ewentualnie „etyki robotów" — *robot ethics* lub *roboethics*)[1] na podstawie trzech książek, które można uznać za reprezentatywne dla obecnej postaci rzeczonego dyskursu: *The New Breed: What Our History with Animals Reveals about Our Future with Robots* autorstwa Kate Darling, *Rights for Robots. Artificial Intelligence, Animal and Environmental Law* popełnionej przez Joshuę Gellersa oraz *Automation and Utopia. Human Flourishing in a World without Work* Johna Danahera.

Ujmując sprawy najogólniej, można uznać, że w ramach robofilozofii i roboetyki główny przedmiot namysłu stanowią zmiany, jakie zachodzą i jakie mogą zajść w wyniku coraz większej obecności coraz bardziej zaawansowanych robotów. W namyśle nad owymi zmianami w szczególności bierze się pod uwagę ich zakres i intensywność oraz ich znaczenie, tudzież ocenę tego znaczenia. Innymi słowy, próbuje się ustalić, w jakich obszarach zachodzą (lub mogą zajść) wywoływane przez rosnącą obecność robotów transformacje, jakie jest (i może być) ich natężenie, oraz czy należy oceniać je (i ich

[1] Termin „robophilosophy" w przyjętym tu znaczeniu został ukonstytuowany przez zespół badawczy kierowany przez Johannę Seibt (Seibt 2018) i stanowi hasło przewodnie odbywających się od 2014 r. co dwa lata konferencji, które gromadzą środowisko skupione wokół dyskursu robofilozoficznego i roboetycznego. Dobre przybliżenie owego dyskursu stanowią publikacje wydawane po każdej z rzeczonych konferencji przez wydawnictwo IOS Press (Seibt et. al. 2014; Seibt et. al. 2016; Coeckelbergh et. al. 2018; Nørskov et. al. 2020), a także — niezwiązane już bezpośrednio z rzeczoną konferencją dwa zbiory artykułów wydanych pod tytułem *Robot ethics* (Lin et. al. 2011; Lin et al. 2017).

konsekwencje) pozytywnie, czy negatywnie. Rzeczony wpływ ujawnia się na dwóch przenikających się, lecz wartych rozróżnienia płaszczyznach. Pierwsza to płaszczyzna spontanicznego doświadczania rzeczywistości. Na jej gruncie najczęściej rozważa się wpływ robotów na dobrostan człowieka wynikający z jego zmieniających się ze względu na obecność robotów doświadczeń w różnych wymiarach życia, np. ekonomicznym, intymnym. Druga płaszczyzna dotyczy refleksyjnego konceptualizowania rzeczywistości. Na jej gruncie rozważa się, w jaki sposób obecność robotów może przyczynić się do zmiany naszego myślenia zarówno o robotach (np. o ich statusie moralnym), jak i o ludziach (w szczególności o ich statusie w porównaniu ze statusem innych bytów) oraz — mówiąc najogólniej — całej rzeczywistości.

Warto podkreślić, że dyskurs robofilozoficzny i roboetyczny zajmuje się robotami, które są „zwykłymi" maszynami w tym sensie, że nie mają cech takich, jak świadomość, emocje, wolna wola itp., lecz co najwyżej mniej lub bardziej przekonująco wiernie zachowania istot ludzkich i (innych) zwierząt. Rozważania dotyczące możliwości uposażenia artefaktów technicznych w rzeczone cechy stanowią raczej marginalną część dyskursu robofilozoficznego i roboetycznego (choć oczywiście stanowią centralny element innych dyskursów związanych z techniką i — w szczególności — sztuczną inteligencją). Ponadto warto zaznaczyć, że choć określenia „robofilozofia" i „roboetyka" sugerują ściśle filozoficzny charakter rzeczonego dyskursu, to w praktyce powszechnie uczestniczą w nim przedstawiciele i przedstawicielki innych dyscyplin z nauk społecznych i humanistyki, takich jak socjologia, prawo, czy antropologia kulturowa, a także programiści i inżynierowie biorący udział w projektowaniu i budowaniu robotów oraz osoby funkcjonujące w ramach interdyscyplinarnego pola badawczego o nazwie „Human-robot interaction" (HRI).

Sformułowane tu ogólne ramy problemowe przywoływanego tutaj dyskursu zostaną dalej wypełnione poprzez rozważania zawarte we wzmiankowanych uprzednio publikacjach książkowych. Wybór tych właśnie książek, spośród wielu innych, podyktowany jest faktem, że prezentują dość zdywersyfikowane stanowiska, dzięki czemu pozwalają przedstawić względnie szerokie spektrum problemów i stanowisk charakterystycznych dla dyskursu, którego część stanowią. Ponadto są to prace opublikowane niedawno, w związku z czym zaznajamiają z aktualnym stanem rzeczonego dyskursu (choć oczywiście wynikają z jego wcześniejszych stadiów i odnoszą się do nich). Zatem, nie przyjmując — ani nawet nie sugerując — ryzykownej tezy, że omówione poniżej prace są najważniejszymi głosami w dyskursie roboetycznym i robofilozoficznym, można jednak uznać, że są one dla owego dyskursu w znacznej mierze reprezentatywne.

Kate Darling z wykształcenia jest prawniczką oraz ekonomistką i zajmuje się problemami związanymi projektowaniem, wdrażaniem i użytkowaniem

robotów. Darling o czym stanowi już sam tytuł jej książki — *The New Breed: What Our History with Animals Reveals about Our Future with Robots* — uznaje roboty za nową rasę, nowy gatunek. Zdaniem autorki, roboty są pod wieloma — choć, rzecz jasna, nie wszystkimi — względami jak zwierzęta, w związku z czym ze zdecydowaną większością problemów, jakie wynikają z rosnącej automatyzacji i robotyzacji mieliśmy już do czynienia w odniesieniu do naszych relacji ze zwierzętami. Tym samym Darling sądzi, że roboty nie wymagają od nas ani zmiany naszego życia, ani naszego myślenia i prognozuje, że sytuacja ta przez dłuższy czas nie ulegnie zmianie, albowiem najbardziej zaawansowane technicznie współczesne roboty są wciąż zaskakująco niedoskonałe i ograniczone, szczególnie w spektrum realizowanych przez nie funkcji. Autorka stanowczo oponuje przeciwko wizjom robotycznej rewolucji, szczególnie tym, które wieszczą rozmaite kryzysy i apokalipsy, zamiast tego proponując, by zastanawianie się nad tym, co mogłoby być (czyli robotycznej rewolucji, której, jej zdaniem, bardzo długo nie będzie) zastąpić namysłem nad tym, co jest (stosunkowo prymitywne roboty), i co było (nasze wielowiekowe współistnienie ze zwierzętami, z którego powinniśmy wyciągać wnioski w odniesieniu do robotów). Zatem główną myśl książki stanowi przekonanie, że jeśli udało się nam wykorzystać zwierzęta dla naszego dobra, to z pewnością jesteśmy również w stanie zrobić to samo z robotami.

Darling rozpoczyna swój wywód od podjęcia dwóch głównych problemów związanych ze stosowaniem robotów do rozmaitych zadań, które określić można mianem technicznych: począwszy od tych obecnych na rynku pracy, a skończywszy na przynależnych do sfery militarnej. Pierwszy z nich to problem technologicznego bezrobocia (Darling 2021, 53-73) podnoszony już przez luddystów u początków industrializacji: jeśli roboty zastąpią ludzi, to ludzie stracą pracę. Darling uważa, że to problem pozorny. Sądzi, że nawet jeśli roboty będą zastępować ludzi, to zawsze pojawiać się będą nowe zawody, chociażby związane utrzymywaniem robotów w należytej kondycji. Podkreśla, że roboty są wciąż bardzo ograniczone i tym samym nie są w stanie całkowicie zastąpić ludzi — mogą ułatwić ludzką pracę i uczynić ją mniej żmudną, ale nie są w stanie — i jeszcze długo nie będą — całkowicie przejąć zadań realizowanych przez istoty ludzkie. Twierdzi także, że nie należy ulegać pesymistycznym wizjom skrajnego determinizmu technologicznego i traktować zastępowania ludzi przez roboty na rynku pracy jako nieuchronnego — ostatecznie to przecież my, ludzie, projektujemy roboty i powinniśmy robić to w taki sposób, by pomagały nam, a nie szkodziły. Zdaniem Darling, z robotami w pracy powinno być — i jej zdaniem będzie — jak z pracą zwierząt, które to zjawisko również obszernie omawia (Darling 2021, 16-52). Koń zatrudniony do ciągnięcia pługa nie zastąpił człowieka, lecz wspomógł go, w wyniku czego praca człowieka stała się lżejsza i efek-

tywniejsza — roboty należy wdrażać tak, by funkcjonowały w sposób analogiczny.

Drugi problem związany z robotami realizującymi głównie zadania techniczne to problem odpowiedzialności za ich funkcjonowanie (Darling 2021, 74-99). Problem ten ujawnia się szczególnie wyraźnie wówczas, gdy mamy do czynienia z maszynami o dużym stopniu autonomiczności, które mogą stanowić zagrożenie, jak np. autonomiczne pojazdy, a także — wspomniane wcześniej — roboty militarne. Kto ma ponieść odpowiedzialność, gdy autonomiczny samochód kogoś potrąci, lub gdy autonomiczny robot militarny ostrzela cywilów zamiast oddziałów wroga? Darling dość krytycznie odnosi się do prób sformułowania teoretycznych, etycznych reguł, które miałyby uniwersalnie rozstrzygać wzmiankowane wyżej kwestie odpowiedzialności. Jej zdaniem wystarczy dotychczasowa praktyka, w szczególności prawna, a konkretnie praktyka przydzielania odpowiedzialności za zachowania zwierząt. Mamy regulacje prawne na wypadek pogryzienia przez psa i analogiczne regulacje zastosować można do funkcjonowania robotów. Darling jest sceptyczna wobec podejść delegujących odpowiedzialność na same roboty — pisze w tym kontekście o średniowiecznych procesach sądowych, gdzie na ławie oskarżonych zasiadały zwierzęta i twierdzi, że współczesne pomysły na obarczanie robotów odpowiedzialnością są równie absurdalne (Darling 2021, 90-94). Zdaniem Darling, jeśli udało się nam rozstrzygnąć problem odpowiedzialności za — z punktu widzenia prawa szalenie skomplikowane — funkcjonowanie pszczół, to bez większych problemów poradzimy sobie również z robotami.

W dalszej kolejności Darling przygląda się możliwości bliskich, emocjonalnych relacji z robotami (Darling 2021, 100-183), w szczególności — lecz nie tylko — tymi zaprojektowanymi do opiekowania się ludźmi lub uprawiania z nimi seksu. Uważa, że w tym zakresie również często mamy do czynienia z nieuzasadnioną „moralną paniką". Darling wymienia liczne przykłady emocjonalnego przywiązywania się do robotów, jak pogrzeby robo-psów AIBO, czy nadawanie imion automatycznym odkurzaczom i odnosi się do obawy, że bliskie relacje z robotami zastąpią bliskie relacje z ludźmi. Uważa, że skoro zwierzęta z reguły nie są substytutami, lecz suplementami ludzi, to nieuzasadnione są obawy, że w przypadku robotów będzie inaczej. Ponadto jeśli nie mamy nic przeciwko hipoterapii, to dlaczego nasze wątpliwości budzą np. roboty stosowane do terapii dzieci w spektrum autyzmu, skoro ich funkcjonowanie nierzadko przynosi pozytywne rezultaty?

Wreszcie Darling podejmuje problem empatii wobec robotów i pomysłów na przyznanie im praw (Darling 2021, 184-234). Przytacza szereg badań empirycznych sugerujących, że zarówno na poziomie świadomych deklaracji, jak i nieuświadamianych reakcji naszego mózgu w pewnej mierze empatyzujemy z robotami i zastanawia się, czy powinniśmy legitymizować to em-

patyczne nastawienie, nadając robotom prawa. Odpowiedź Darling brzmi „tak", choć podkreśla ona, że prawa te powinniśmy nadać nie ze względu na roboty, lecz ze względu na nas samych, ludzi. Nie chcielibyśmy, twierdzi autorka, by ludzie krzywdzili swoje roboty na oczach dzieci (nawet jeśli roboty będą jedynie symulować cierpienie, a nie faktycznie go doświadczać, podobnie jak w przypadku filmów fabularnych przedstawiających przemoc wobec ludzi i ich cierpienie) — nadajmy więc tym robotom prawa, by chronić nie tyle je, ale właśnie nasze dzieci i nas samych. Darling twierdzi, że podobnie jak w przypadku ludzkich zwłok, aktualnie istniejącym robotom (nieposiadającym świadomości, zdolności odczuwania cierpienia itp.) nie będziemy przyznawać praw ze względu na to, czym są i jakie cechy mają, ale ze względu na nasze emocje względem nich.

Podsumowując, Darling stara się dowieść, iż technika w ogóle, a roboty w szczególności, są narzędziami mającymi wzmacniać ludzki dobrostan, w związku z czym nie należy budować na ich temat obszernych teorii, lecz wdrażać efektywne regulacje praktyczne. Takie antropocentryczne i praktycystyczne stanowisko, które mocno miarkuje nasze zachwyty i obawy dotyczące technicznych rewolucji, jest stosunkowo częste — od dłuższego czasu prezentuje je choćby — nieco bardziej radykalna w swoim antropocentryzmie i praktycyzmie — Joanna J. Bryson (2018), a wśród niedawno wydanych książek, można je znaleźć choćby w dociekaniach wpływowego filozofa techniki Svena Nyholma (2020), który co prawda prezentuje podejście bardziej abstrakcyjne i spekulatywne, ale ostatecznie zajmuje stanowisko bardzo zbliżone do Darling.

Mimo że Darling słusznie odczarowuje nasze myślenie o technice, wskazując, że nie jest ona czymś tak rewolucyjnym i tajemniczym jak czasem skłonni jesteśmy sądzić, to wydaje się, że niejednokrotnie idzie zbyt daleko jeśli idzie o optymizm, upraszczanie i stawianie na analogię pomiędzy robotami i zwierzętami. Autorka zdaje się zapominać, że dotychczasowy proces automatyzacji i zastępowania człowieka przez maszyny na rynku pracy bynajmniej nie był pozbawiony tzw. kosztów ludzkich, nie wspominając o degradacji środowiska — w tym kontekście trudno zatem po prostu zgodzić się z jej tezą, że dalsza, najpewniej szersza zakresowo i bardziej intensywna niż dotychczas automatyzacja zapewni „miękkie lądowanie" wszystkim, którzy utracą w jej wyniku źródło utrzymania, a jednocześnie będzie neutralna dla środowiska — ponadto istnieje wiele stanowisk głoszących, iż owa automatyzacja doprowadzi do znacznie głębszych zmian, niż sugeruje Darling. Kolejne zastrzeżenia wobec stanowiska Darling wynikają z faktu, że porównując relacje z robotami do relacji ze zwierzętami, Darling unika szeregu trudnych i istotnych problemów. Ignoruje np., że w kontekście swojej funkcji w ramach relacji seksualnych roboty zdają się mieć więcej wspólnego z pornografią niż

ze zwierzętami, natomiast wpływ tej ostatniej na społeczny dobrostan jest — jak wiadomo — wysoce kontrowersyjny (Sparrow 2017). Darling ignoruje ten problem i bardzo obszerną dyskusję na temat negatywnego wpływu relacji z robotami w postaci np. intensyfikacji zjawiska uprzedmiotowienia kobiet (Richardson 2016) lub erozji relacji międzyludzkich (Turkle 2012) na ten temat jako wzniecanie niepotrzebnej paniki.

Zupełnie inaczej postrzega sprawy Joshua Gellers, który jest politologiem, ale zajmuje się również prawem. Gellers skupia się na kwestii praw robotów, co dość wyraziście sygnalizuje tytuł jego książki: *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*. Również przywołuje przy tej okazji prawa zwierząt, ale w ostatecznym rozrachunku uważa je za nie w pełni satysfakcjonujący punkt odniesienia. Wynika to z faktu, że — w przeciwieństwie do Darling — Gellers chce uniknąć antropocentrycznego przyznawania robotom praw ze względu na ludzi (podobnie jak, jego zdaniem, ma to zwykle miejsce w odniesieniu do zwierząt), lecz proponuje wdrożyć owe prawa na sposób antyantropocentryczny, posthumanistyczny. Tym samym Gellers nie chce po prostu nadać robotom praw chroniących interesy ludzi, a jednocześnie pozostawiających im status — co prawda szczególnych, bo obdarzonych prawami, ale wciąż — przedmiotów, rzeczy. W tym sensie, o ile dla Darling roboty stanowią praktyczne wyzwanie, któremu w znacznej mierze podołaliśmy w przypadku zwierząt, o tyle dla Gellersa stoimy przed wyzwaniem nie tylko praktycznym, ale również teoretycznym, wyzwaniem zmiany naszego myślenia nie tylko o robotach, czy zwierzętach, ale o nas samych i naszej ontologii, tak, by w ostatecznym rozrachunku stała się ona — *nomen omen* — mniej nasza, mniej antropocentryczna.

Gellers bynajmniej nie stroni od empirycznego konkretu, dość często przywołując również konkretne casusy prawne występujące w rozmaitych systemach prawnych i dotyczące nadawania praw nieludziom, jednak w porównaniu z Darling oferuje znacznie więcej teorii i abstrakcji. Jako główne źródła swoich inspiracji wymienia tematykę antropocenu w filozofii, zwrot materialistyczny w humanistyce i naukach społecznych, zwrot ontologiczny w prawie środowiskowym i zwrot relacyjny w etyce oraz myśleniu o robotach (Gellers 2021, 9). Tym samym Gellers odwołuje się do nurtów takich jak ekologia głęboka czy nowy materializm, ale także do myśli niezachodniej i nienowoczesnej, jak światopoglądy ludów tubylczych czy buddyzm. Jeśli dodamy do niego wspomniane już odniesienia do różnych — w tym również niezachodnich — systemów prawnych i sposobu, w jakie ujmują one prawa nieludzi, otrzymujemy imponujący tygiel konceptualny, który jednocześnie stanowi wyraz narastającej w robofilozoficznym i roboetycznym dyskursie refleksji, by sposobów konceptualizacji problemów związanych z robotami poszukiwać poza myślą nowoczesną i zachodnią.

By uczynić rzeczony tygiel znośnym dla czytelnika, Gellers rozpoczyna od uporządkowania kilku podstawowych pojęć i stanowisk istotnych dla dyskursu robofilozoficznego i roboetycznego (Gellers 2021, 15-61). Wskazuje na rozmaite podejścia do praw robotów, wydzielając przede wszystkim te, które postulują, by przyznawać robotom prawa ze względu na ich cechy (np. świadomość, czy zdolność odczuwania cierpienia) oraz te o charakterze relacyjnym, w przypadku których najważniejsze są nie tyle cechy robotów, co sposób wchodzenia z nimi w interakcje (czyli na przykład to, jakie emocje budzą w nas roboty) (Gellers 2021, 16-23). Rozróżnia także i drobiazgowo omawia dwie koncepcje przyznawania praw: na podstawie woli danego podmiotu lub jego interesów (Gellers 2021, 50-53). Gellers analizuje też różne rodzaje osobowości — od osobowości moralnej (etycznej) poczynając, przez osobowość psychologiczną, a na osobowości prawnej skończywszy (Gellers 2021, 28-42). Szczególną aprobatę Gellersa budzi koncept osobowości relacyjnej, którego respektowanie przypisuje on ludom tubylczym.

Gellers poświęca obszerny wywód problemowi praw zwierząt, szczegółowo omawiając podejście do niego na gruncie rozmaitych religii i nurtów myślowych (Gellers 2021, 62-103). Dochodzi do wniosku, że jeśli nawet w zamierzeniu zwolenników praw zwierząt owe prawa mają stanowić przekroczenie antropocentryzmu, to w większości przypadków pozostają one pod wieloma względami antropocentryczne. Rozstrzygające są kryteria ustalane przez ludzi, którymi są, albo posiadanie przez zwierzęta cech wystarczająco podobnych do cech ludzkich, albo wywoływanie w ludziach określonych reakcji emocjonalnych, nie wspominając o podejściach zbliżonych do stanowiska Darling odnośnie do praw robotów, gdzie nadaje się zwierzętom prawa nie tyle ze względu na dobro ich samych, co ze względu na dobro ludzi. I mimo, że Gellers uznaje, że pewien stopień antropocentryzmu jest w zasadzie niemożliwy do uniknięcia, a także dostrzega zalety niektórych pomysłów pojawiających się na gruncie myślenia o prawach zwierząt (np. przyznanie zwierzętom praw własności do terenu, który zamieszkują), to w ostatecznym rozrachunku uznaje, iż prawa zwierząt nie stanowią dobrej inspiracji do właściwego rozstrzygnięcia kwestii praw robotów.

Z nieco większym entuzjazmem odnosi się Gellers do idei praw natury (Gellers 2021, 104-139). W szczególności afirmuje obecny w myśleniu o nich holizm, przekonanie o potrzebie całościowego myślenia o przyrodzie, zamiast indywidualistycznego pochylania się nad konkretnymi gatunkami obecnego w myśleniu o prawach zwierząt. Jednak również ten rodzaj myślenia jest zdaniem Gellersa niewystarczająco emancypacyjny i egalitarny. Dlaczego bowiem mielibyśmy przyznawać prawa i tym samym chronić tylko naturę, przyrodę? Podążając za licznymi inspiracjami, które najogólniej można określić mianem posthumanistycznych, Gellers postuluje uczynienie

holizmu skrajnie synkretycznym, odrzucenie platońskich, kartezjańskich i wszelkich innych dualizmów, zatarcie granic między naturą i kulturą, ożywionym i nieożywionym, ludzkim i nieludzkim. Sądzi, że na gruncie takiej „płaskiej", pozbawionej hierarchii ontologii uda się zbudować wreszcie podejście acentryczne, uniknąć nie tylko antropocentryzmu, ale również np. biocentryzmu, tym samym rezygnując z nobilitacji jakiegokolwiek rodzaju bytu.

Rozważania Gellersa wpisują się w głęboko już zakorzeniony w humanistyce i naukach społecznych nurt posthumanizmu, coraz szerzej obecny również w namyśle nad robotami i techniką. Gellers staje zatem w szeregu wspólnie z również debatującym na temat praw robotów jako ucieczki od antropocentryzmu Davidem Gunkelem (2018), czy wykorzystującym myśl posthumanistyczną do uzasadnienia konieczności podmiotowego traktowania robotów Danielem Estradą (2020). W tym sensie praca Gellersa jest z całą pewnością znakomitym ćwiczeniem nieatropocentrycznego myślenia o świecie, a w szczególności o robotach i stanowi jeden z najpełniejszych wyrazów tego stanowiska na gruncie dyskursu robofilozoficznego i roboetycznego.

Niemniej antyantropocentryczne, posthumanistyczne stanowisko Gellersa, Gunkela i innych spotyka się ze stanowczą krytyką ze strony bardziej tradycyjnie zorientowanych badaczy i badaczek. Po pierwsze, krytyka ta ma charakter aksjologiczny i negatywnie odnosi się do debatowania na temat praw robotów i troski o artefakty techniczne w sytuacji, gdy łamane są prawa ludzi, nierzadko w trakcie procesów produkcyjnych rozmaitych technicznych artefaktów. W ramach tego typu krytyk twierdzi się, że debaty na temat praw robotów to niepoważna rozrywka filozofów ignorujących prawdziwe problemy dotyczące krzywd doznawanych przez ludzi, a także wyraża się przekonanie, że nadanie praw robotom w żaden sposób nie poprawi ani dobrostanu robotów (ponieważ nie ma sensu mówienie o czymś takim jak dobrostan robotów), ani dobrostanu — nierzadko ignorowanych w posthumanistycznych debatach na temat praw robotów — ludzi (Birhane, van Dijk 2020). Po drugie, krytyka ta ma charakter epistemologiczny i wskazuje na fakt, iż osoby postulujące przyznanie praw robotom nie potrafią wskazać poznawczego uzasadnienia tego postulatu lub formułują uzasadnienia całkowicie nieprzekonujące. W szczególności idzie o to, że osoby postulujące przyznanie praw robotom traktują jako nieistotne to, że roboty te nie mają cech takich, jak świadomość, zdolność odczuwania cierpienia itp. Zakładają bowiem sceptycznie, że faktu występowania tych cech nie da się w pełni wiarygodnie ustalić nie tylko w przypadku robotów, ale również ludzi i zwierząt, w związku z czym uprawnione jest przyjęcie innych uzasadnień przy nadawaniu rzeczonych praw, które z kolei przeciwnicy nadawania praw robotom postrzegają jako arbitralne i irracjonalne (Müller 2021).

Darling i Gellers prezentują dość szerokie spektrum myślenia o robotach i technice, choć oczywiście są dalecy od jego wyczerpania. By się o tym przekonać, warto pochylić się nad twórczością Johna Danahera. Ten irlandzki filozof techniki i prawnik w książce *Automation and Utopia. Human Flourishing in a World without Work* wychodzi od jednego z problemów, którymi zajmuje się Darling: zastępowania ludzi przez roboty na rynku pracy. Sądzi, że nieuchronnie nadchodzi „jesień człowieczeństwa" i wkrótce zamiast w antropocenie, będziemy żyć w robotocenie (Danaher 2019, 2). Danaher nie zgadza się zatem z Darling, że roboty niewiele zmienią w naszym życiu, ponieważ sądzi, że automatyzacja zupełnie zmieni nasz świat i jego doświadczanie. Nie zgadza się też jednak również z Gellersem, który postuluje zmianę naszego myślenia na mniej antropocentryczne, jako że centralną wartością jest dla niego sens ludzkiego życia.

Danaher, w odróżnieniu od Darling, jest głęboko przekonany, iż roboty zastąpią ludzi w większości realizowanych przez nich aktywności, w szczególności na rynku pracy, a tym samym sprawią, że będziemy mieli do czynienia z masowym technologicznym bezrobociem i zdecydowana większość ludzi na świecie nie będzie mogła znaleźć pracy (Danaher 2019, 25-52). Jednakowoż ekonomiczne konsekwencje tego scenariusza nie są, zdaniem filozofa, powodem do obaw. Danaher uważa bowiem, że powszechna automatyzacja spowoduje powszechny dobrobyt: maszyny będą wytwarzać mnóstwo dóbr, natomiast kapitalizm zostanie zreformowany w taki sposób, iż wdrożona zostanie efektywna redystrybucja sprawiająca, iż wszyscy będą mieli do rzeczonych dóbr dostęp — w tym ostatnim kontekście przytacza rozmaite konkretne rozwiązania, by wspomnieć choćby o bezwarunkowym dochodzie podstawowym.

Obawy Danahera budzą natomiast zagadnienia egzystencjalne: uważa bowiem, że w wyniku nadmiernej automatyzacji ulegnie erozji sens ludzkiego życia (Danaher 2019, 87-131). Po pierwsze, automatyzacja sprawi, że utracimy swoją sprawczość, że przestaniemy mieć poczucie wpływu na otaczający świat i satysfakcję z kształtowania go na własną modłę. Świat będą zmieniać maszyny, które programować będzie wąska grupa osób, natomiast wszyscy pozostali będą jedynie biernie konsumować płynące z tych zmian benefity. Po wtóre, nie tylko nie będziemy potrafili świata zmieniać, ale przestaniemy go również rozumieć: proces automatyzacji sprawi, że przestaniemy intelektualnie ogarniać rzeczywistość, nadążać za ingerencjami dokonywanymi przez maszyny i interpretować dane, jakie na temat owego świata pozyskują. Co więcej, możemy w ogóle przestać cokolwiek rozumieć, a to ze względu na fakt, iż technika utrudnia nam zdolność koncentracji na ważnych sprawach, w najlepszym wypadku koncentrując ją na mało istotnych, błahych kwestiach, a w najgorszym permanentnie nas rozpraszając poprzez podsuwanie coraz

to nowych źródeł wrażeń. W rezultacie, tracąc działaniowy wpływ na rzeczywistość, a także intelektualne zdolności jego rozumienia, w istocie utracimy swoją autonomię — w swoim zdemotywowaniu i zdezorientowaniu będziemy robili i myśleli to, co podsuwa nam pod nos technika.

Tę dystopijną wizję Danaher stara się zastąpić wizją utopijną. Po szczegółowych rozważaniach dotyczących samego pojęcia utopii (Danaher 2019, 135-156) myśliciel proponuje dwie utopijne wizje: cyborgiczną (Danaher 2019, 157-213) i wirtualną (Danaher 2019, 214-270). Już z samych nazw tych utopii można odczytać kolejną zasadniczą różnicę między Danaherem a Darling i Gellersem. Podczas gdy wspomniana dwójka szuka sposobów na radzenie sobie z techniką w myśleniu o przyrodzie: w naszym myśleniu o zwierzętach i środowisku naturalnym, Danaher szuka rozwiązań technicznych: lekiem na nadmiar techniki ma być jeszcze więcej techniki.

Cyborgiczna utopia polega na spełnieniu transhumanistycznych snów o zintegrowaniu się z techniką: jeśli technika nas przerasta, to sprawmy, by stała się częścią nas. Danaher sądzi, że w ten sposób będziemy w stanie konkurować z maszynami na każdym polu, a co za tym idzie nie oddamy im sprawczości i rozumienia: nadal będziemy mogli efektywnie zmieniać świat na naszą modłę i ogarniać go intelektualnie. Ponadto Danaher zauważa, iż cyborgizacja człowieka przybliża nas do kolonizacji innych planet (Danaher 2019, 178-184). Z jednej strony, twierdzi filozof, jest to wyzwanie, które samo w sobie pomaga nadać naszym życiom sens, a z drugiej strony, zasiedlanie innych planet pozwoliłoby zapewnić naszemu gatunkowi dłuższe trwanie — gdyby jedna z planet przestała nadawać się do zamieszkania, mielibyśmy inne. Ostatecznie jednak Danaher uznaje cyborgiczną utopię za niezadowalającą. Pośród licznych i frapujących argumentów krytycznych szalę zdają się przeważać te praktyczne: cyborgizacja człowieka rozwija się zbyt wolno, by mogła łagodzić skutki galopującej automatyzacji, a ponadto organizm ludzki jest zbyt mało rozpoznany, by mieć pewność, że cyborgiczne ingerencje nie będą wywoływać negatywnych skutków ubocznych.

Znacznie większy entuzjazm budzi w Danaherze utopia wirtualna, czyli przeniesienie przynajmniej części naszej egzystencji do światów wirtualnych. Danaher zaczyna od próby odparcia narzucającego się wręcz zarzutu, iż to, co wirtualne, nie jest realne, poprzez próbę zatarcia różnicy między realnością i wirtualnością przekonując, że na co dzień doświadczamy wielu wirtualnych rzeczy jako realnych (zwykle nazywając je symbolicznymi), więc podobnie może być z wirtualną rzeczywistością (Danaher 2019, 217-230). Danaher przekonuje, iż wirtualna rzeczywistość pozwoli nam na wzmocnienie poczucia sensu życia, ponieważ będziemy mieli wpływ na jej kształt (jak np. w Minecrafcie), a także dostarczy nam możliwości pozyskiwania osiągnięć, ze względu na obecną w owej wirtualności rywalizację w ramach rozmaitych gier — rzec

można, iż luddyczne obawy dotyczące techniki Danaher zamienia na ludyczny entuzjazm (Danaher 2019, 231-251). Kolejną, zdaniem filozofa, zaletą wirtualnej utopii jest fakt, iż nie musi się ona ograniczać do jednego wirtualnego świata. Takich światów może być wiele — mogą być od siebie w znacznej mierze niezależne, prezentować różne reguły, dostarczać różnych możliwości i — co najważniejsze — oferować różne wizje dobrego życia (Danaher 2019, 251-269). W tym sensie Danaher jest przekonany, że gdy — ze względu na automatyzację — nasz materialny świat przestanie stanowić dla nas źródło sensu, powinniśmy wykorzystać możliwości, jakie dają nam światy wirtualne, a — wedle filozofa — możliwości te są w zasadzie nieograniczone.

Książka Danahera wpisuje się w przynajmniej dwa pola problemowe. Po pierwsze, stanowi głos w bardzo obszernej w ostatnich latach dyskusji na temat automatyzacji i technologicznego bezrobocia, w ramach której jedno z bardziej wpływowych stanowisk stanowią prace autorstwa Daniela Susskinda, w tym jego najnowsza książka (Susskind 2020). Po drugie, należy do nurtu, który prognozuje poważne problemy związane z rozwojem techniki, a zarazem to właśnie w technice dostrzega możliwość rozwiązania owych problemów: podobne stanowisko odnaleźć można choćby u Juliana Savulescu i Ingmara Perssona (Savulescu, Persson 2012), którzy sądzą, że stworzyliśmy świat, nad którym — m.in. ze względu na rozwój techniki — nie jesteśmy w stanie zapanować, a rozwiązaniem tego stanu rzeczy miałoby być sztuczne (tzn. polegające na ingerencji genetycznej i/lub farmakologicznej) ulepszanie ludzkiej moralności.

Mimo że propozycja Danahera ma charakter antropocentryczny i dobro człowieka stanowi w niej wartość centralną, to właśnie obecne w niej rozumienie człowieka może pod pewnymi względami zastanawiać. Danaher wydaje się bowiem pomijać wartość tego, co często uznaje się za typowo ludzkie. Gdy zapowiada świat galopującej automatyzacji, właściwie nie wskazuje na dość oczywisty fakt, że pozbawieni konieczności pracy zarobkowej mielibyśmy znacznie więcej czasu na relacje miłosne, seksualne, rodzinne i przyjacielskie z bliskimi nam ludźmi (i nieludźmi) i to w nich odnajdywać sens życia. Pierwsza przyczyna tego stanu rzeczy może wynikać z faktu, że Danaher nie uznaje takich emocjonalnych relacji z innymi osobami za szczególnie istotne, a za znacznie bardziej relewantne uważa poznawcze i twórcze aktywności istot ludzkich. Druga przyczyna znikomej uwagi poświęconej bliskim relacjom może brać się stąd, że zdaniem Danahera również na gruncie takich bliskich relacji nastąpi automatyzacja i robotyzacja, ludzie zostaną zastąpieni przez roboty. W rzeczy samej irlandzki filozof jest zwolennikiem poglądu określanego mianem etycznego behawioryzmu, stanowiącego (w najbardziej uproszczonym ujęciu), że jeśli zachowanie danego robota wystarczająco przypomina zachowanie człowieka, to owego robota należy traktować

właśnie jak człowieka, co z kolei prowadzi go do przekonania, iż możliwa jest autentyczna przyjaźń z robotem (nawet takim, który nie ma uczuć ani świadomości) (w kwestii gruntownej krytyki poglądów Danahera zob. Smids 2020). Ponadto propozycja funkcjonowania w wirtualnej rzeczywistości i zastąpienia ciała wirtualnym awatarem — mimo iż ma po swojej stronie szereg sojuszników, jak np. zespół pod kierunkiem Luciana Floridiego twierdzący, że rozróżnienie na rzeczywistość online i offline nie ma już sensu, ponieważ obu tych rzeczywistości doświadczamy na tyle podobnie, że w gruncie rzeczy stanowią jedną, homogeniczną rzeczywistość (Floridi 2015) — wciąż jawi się jako wysoce kontrowersyjna. Ignoruje bowiem pewne różnice pomiędzy rzeczywistością materialną i wirtualną (np. kwestię dotyku), a także wymaga daleko idącej zmiany naszego myślenia o tym, jaki jest status ontologiczny rzeczywistości wirtualnej. Powyższe sygnały sugerują — a potwierdzają to w znacznej mierze inne, w tym wspomniane wyżej, publikacje autora — iż Danaher w swoisty sposób „odczarowuje człowieka", nie dostrzegając w nim niczego wyjątkowego i specyficznie ludzkiego, a tym samym nie tyle przedstawiając roboty jako coraz bardziej podobne do człowieka, co raczej ujmując człowieka jako właściwie nieróżniącego się w żaden istotny sposób od robotów. Mimo że rezultat tej propozycji jest podobny do tego proponowanego przez Gellersa — człowiek nie jest wyróżniony spośród innych bytów — to jednak drogi dojścia do tego rezultatu są odmienne. Danaher zajmuje stanowisko naturalistyczne, tudzież redukcjonistyczne, postrzegając człowieka jako biologiczną maszynę, natomiast Gellers w duchu posthumanizmu stara się nie tyle zredukować człowieka do poziomu innych bytów, co przede wszystkim nadać innym bytom status analogiczny do ludzkiego.

Podsumowując, warto wypunktować centralne problemy i stanowiska charakterystyczne dla dyskursu robofilozoficznego i roboetycznego ujawniające się w omówionych tu pozycjach. Po pierwsze, problem stanowi diagnoza i prognoza zakresu i intensywności zmian związanych z narastającą obecnością robotów. W ramach stanowiska, które reprezentuje m.in. Darling, zmiany te będą stosunkowo niewielkie i będą mieć charakter ewolucyjny, natomiast w ujęciu Danahera i Gellersa owe zmiany będą mieć charakter głęboki i radykalny, zarówno jeśli idzie o spontaniczne, praktyczne doświadczanie rzeczywistości, jak i jej refleksyjną, teoretyczną konceptualizację. Rozpiętość owych diagnoz ujawnia się w przypadku omawianych tu prognoz dotyczących automatyzacji oraz praw robotów. Jednym z czynników, które warunkują tak dużą rozpiętość wzmiankowanych wyżej stanowisk, jest fakt, że Darling skłania się ku instrumentalizmowi technologicznemu. Uważa, że technika jest jedynie narzędziem, którego ludzie używają jako środka do swoich celów, i że będziemy w stanie dostosować technikę do siebie, podczas gdy Danaher i Gellers zdają się przejawiać pewne cechy determinizmu technologicznego zakładają-

cego, że rozwój techniki nie tylko nie jest w pełni zależny od stawianych przez ludzi celów, ale również w pewnej mierze owe cele determinuje. Oznacza, że to my musimy dostosować się do techniki. Po drugie, problem zmian obejmuje również ich ocenę. Ocena ta zależy, rzecz jasna, od przyjmowanych stanowisk aksjologicznych, w ramach których wyróżnić można przede wszystkim stanowisko tradycyjnie antropocentryczne oraz posthumanistyczne stanowisko antyantropocentryczne, a także stanowisk epistemologicznych, które można najogólniej podzielić na trzymające się dotychczasowych sposobów konceptualizowania rzeczywistości, oraz takie, które proponują nowe sposoby ujmowania otaczającego nas świata. Rozdźwięk pomiędzy tymi stanowiskami łatwo dostrzec w omawianej tu kwestii oceny bliskich relacji człowieka z robotami, a także sposobów myślenia o statusie moralnym (w tym: prawach) robotów, oraz o statusie ontologicznym człowieka.

Nietrudno dostrzec, że znaczna część przedstawionych tu problemów i stanowisk składających się na dyskurs robofilozoficzny i roboetyczny bynajmniej nie jest specyficzna wyłącznie dla rzeczonego dyskursu, lecz ujawnia się również w przypadku innych dyskusji toczonych na gruncie nauk humanistycznych i społecznych, w tym filozofii. W tym sensie dyskurs robofilozoficzny i roboetyczny stanowi integralną część współczesnej kultury zachodniej i wyraz obecnych w niej tendencji.

Literatura

Birhane, Abeba; van Dijk Jelle (2020). "Robot Rights? Let's Talk about Human Welfare Instead", [w:] *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, s. 207-213.

Bryson, Joanna J. (2018). "Patiency is not a virtue: the design of intelligent systems and systems of ethics". *Ethics and Information Technology*, vol. 20, s. 5-26.

Coeckelbergh, Mark; Loh, Janina; Funk, Michael; Seibt, Johanna; Nørskov Marco (eds.) (2018). *Envisioning Robots in Society — Power, Politics, and Public Space, Proceedings of Robophilosophy 2018*. IOS Press.

Danaher, John (2019). *Automation and Utopia: Human Flourishing in a World without Work*. Harvard University Press.

Darling, Kate (2021). *The New Breed: What Our History with Animals Reveals about Our Future with Robots*. Henry Holt.

Estrada, Daniel (2020). "Human supremacy as posthuman risk". *Journal of Sociotechnical Critique*, vol. 1, no. 1, s. 1-40.

Gellers, Joshua (2021). *Rights for Robots: Artificial Intelligence, Animal and Environmental Law*. Routledge.

Gunkel, David J. (2018). *Robot Rights*. MIT Press.

Floridi, Luciano (ed.) (2015) *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Springer.

Müller, Vincent C. (2021). "Is it time for robot rights? Moral status in artificial entities". *Ethics Information Technology*, vol. 23, s. 579-587.
Nørskov, Marco; Seibt, Johanna; Quick, Oliver Santiago (eds.) (2020). *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy 2020*, IOS Press.
Nyholm, Sven (2020). *Humans and Robots: Ethics, Agency, and Anthropomorphism*. Rowman & Littlefield.
Lin, Patrick; Abney, Keith; Bekey George A. (eds.) (2011). *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press.
Lin, Patrick; Abney, Keith; Jenkins Ryan (eds.) (2017). *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press.
Richardson, Kathleen (2016). "Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines". *IEEE Technology and Society Magazine*, vol. *35*, no. 2, s. 46-53.
Savulescu, Julian; Persson, Ingmar (2012). *Unfit for the Future: The Need for Moral Enhancement*, Oxford University Press.
Seibt, Johanna (2018). "Robophilosophy", w: Braidotti, Rosi; Hlavajova, Maria (eds.). *Posthuman Glossary*. Bloomsbury, s. 390-393.
Seibt, Johanna; Hakli, Raul; Nørskov Marco (eds.) (2014). *Social Robots and the Future of Social Relations. Proceedings of Robophilosophy 2014*. IOS Press.
Seibt, Johanna; Nørskov, Marco; Andersen, Soren Schack (eds.) (2016). *What Social Robots Can and Should Do. Proceedings of Robophilosophy 2016*. IOS Press.
Smids, Jilles (2020). "Danaher's Ethical Behaviourism: An Adequate Guide to Assessing the Moral Status of a Robot?". *Science and Engineering Ethics*, vol. 26, no. 5, s. 2849-2866.
Sparrow, Robert (2017). "Robots, rape and representation". *International Journal of Social Robotics*, vol. 9, no. 4, s. 465-477.
Susskind, Daniel (2020). *A World Without Work: Technology, Automation, and How We Should Respond*, Allen Lane.
Turkle, Sherry (2012). *Alone Together: Why We Expect More from Technology and Less from Each Other.* Basic Books.

Maciej Musiał

Selected Problems and Positions on the Basis of Robophilosophy and Robotics

Abstract

The purpose of this text is to discuss the main problems and positions present within one of the discourses dealing with the issue of robots, which is most often described as "robophilosophy" and "robot ethics." In particular, the paper highlights the problems and positions that refer to the current and future impact of the growing presence of robots and attempts to diagnose and predict the range, intensity and significance of this impact, as well as its evaluation.

Keywords: robophilosophy, robot ethics, robot rights, technological unemployment, automation.