



DAMIAN SZCZĘCH

<https://orcid.org/0000-0003-2290-1726>

Katolicki Uniwersytet Lubelski Jana Pawła II, Wydział Filozofii

W kierunku bezpiecznej przyszłości: etyczne i prawne problemy związane z pojazdami autonomicznymi

Aiming for the safe future: ethical and legal problems associated with autonomous vehicles

ABSTRAKT: Jednym ze sposobów zwiększenia bezpieczeństwa na drogach może być eliminacja czynnika ludzkiego, będącego przyczyną większości wypadków. Jednak co jakiś czas słyszymy o zdarzeniach z udziałem pojazdów autonomicznych. Rozwój technologii zwiększa bezpieczeństwo, lecz także stwarza nowe zagrożenia. Celem niniejszego artykułu jest wskazanie problemów związanych z oceną moralną i prawną negatywnych skutków działania pojazdów autonomicznych oraz przedstawienie propozycji rozwiązań. Ocenę utrudnia fakt, że pojęcie autonomii wiąże się z przyjęciem jakiegoś rodzaju podmiotowości. Proponowane rozwiązanie zakłada wprowadzenie „kontraktu AI” oraz utworzenie funduszu ubezpieczeniowego, którego celem byłoby pokrycie przynajmniej części wyrządzonych szkód – niezależnie od oceny stopnia odpowiedzialności producentów takich pojazdów. Kwestią do rozstrzygnięcia pozostaje natomiast, czy faktycznie będzie można uznać robota za (współ)winnego wyrządzonej szkody.

SŁOWA KLUCZOWE: AI, sztuczna inteligencja, pojazdy autonomiczne, „kontrakt AI”, etyka AI

ABSTRACT: One of the ways increasing road safety might be the elimination of the human factor, which is the cause of most accidents. However, once in a while we hear about incidents involving autonomous vehicles. The development of technology increases safety, but also creates new risks. The purpose of this article is to point out problems related to the moral and legal evaluation of the consequences of using autonomous vehicles, and to suggest solutions. The assessment is made more difficult by the fact that the concept of autonomy implies some kind of subjectivity. The solution in question involves introducing an “AI contract”, and creating an insurance fund the aim of which would be to cover at least part of caused damage – regardless of the assessment of the degree of liability of manufacturers producing such cars. However, the question remains whether it will be possible to consider the robot as accessory to the damage caused.

KEYWORDS: AI, artificial intelligence, autonomous vehicles, “AI contract”, AI ethics

1. Wstęp

Niemal każdego tygodnia słyszymy informacje dotyczące przełomu w rozwoju sztucznej inteligencji (AI), pojawienia się kolejnego modelu językowego i zapowiedzi nadchodzących rozwiązań. Presja i tempo rozwoju są ogromne, ponieważ w świecie nauki i techniki nie ma drugich miejsc – liczy się ten, kto osiągnął dane rozwiązanie jako pierwszy. Sprawia to, że legislacja nie nadąza za technologią, a luka wiedzy między twórcami technologii a organami tworzącymi regulacje prawne jest ogromna¹. Między innymi z tych powodów istnieje wiele niejasności etycznych i prawnych co do zastosowania rozwiązań opartych na AI, z którymi niemal codziennie wchodzimy w interakcje, nawet jeśli nie jesteśmy tego świadomi. Filtrowanie spamu w wiadomościach mailowych, kolejność linków w wyszukiwarce Google, polecane filmy na Netflix, rekomendacje produktów w sklepach internetowych, systemy wspomagania kierowcy, autonomiczne pojazdy – to tylko niektóre przykłady zastosowania AI.

Technika wpływa na nasze funkcjonowanie i sprawia, że dostosowujemy się do jej sposobu działania. W efekcie, nasze życie bywa organizowane przez coś, co zaczynamy traktować nie tylko jako wytwór człowieka, ale także jako uczestnika życia społecznego. Takie tendencje zostają wzmocnione poprzez dążenie do usamodzielnienia systemów opartych na AI. W rezultacie, może to powodować zrzeczenie się części odpowiedzialności zgodnie z hasłem: „to nie my, to AI tak decyduje”. Usamodzielnienie sprawia także, iż wzrasta zaufanie do technologii, której działania nikt nie rozumie – jest to tzw. problem czarnej skrzynki, który jest kluczowy dla zagadnień poruszanych w niniejszym artykule. Wyjaśnienie sposobu, w jaki AI podejmuje decyzje, jest istotną kwestią etyczną. Zastosowanie takiej technologii w zadaniach, które mogą doprowadzić do utraty zdrowia lub życia innych uczestników ruchu drogowego, jest wątpliwe pod względem moralnym i prawnym. Kwestia odpowiedzialności za ewentualne szkody nadal nie została rozstrzygnięta. Dodatkowymi komplikacjami są niejednoznaczny i nierzadko mylący marketing oraz nieświadomość użytkowników co do możliwości i ograniczeń technologii.

W dyskusjach wokół etyki nowych technologii często zwraca się uwagę na wartości, jakimi powinna się kierować AI w swoich działaniach, oraz na to, że te wartości powinny być zbieżne z naszymi². Warto w tym miejscu podkreślić występowanie silnej zależności między wartościami a kontekstem kulturowym. Oznacza to, że zasady postępowania (także moralnego) są uwarunkowane tery-

¹ C. Campbell, *AI by Design: A Plan for Living with Artificial Intelligence*, CRC Press, New York 2022, s. 51–62.

² N. Savage, *How Robots Can Learn to Follow a Moral Code*, Nature, 26.10.2023, <https://www.nature.com/articles/d41586-023-03258-1> [dostęp: 30.05.2024].

torialnie. Co za tym idzie, AI powinna te różnice uwzględniać w swoim funkcjonowaniu, aby uniknąć problemów związanych z neokolonializmem, którego przejawem jest narzucanie przez wytwory techniki wartości obcych dla danej kultury. Niestety, twórcy rozwiązań technicznych rzadko interesują się humanistyką, przez co takie problemy są im obce. Naszym głównym orężem jest dążenie do prawnych regulacji mających na celu ochronę społecznego interesu, nie tylko w aspekcie ochrony wartości i kultur. Projektując regulacje, nie powinniśmy się skupiać na tym, co AI może zrobić, lecz na tym, co powinna – jej zastosowania powinny dotyczyć wyłącznie obszarów o potencjalnie niedużym ryzyku wyrządzenia szkód.

Głównym celem niniejszego tekstu jest udzielenie odpowiedzi na pytanie: jakie problemy wiążą się z oceną moralną i prawną negatywnych skutków działania pojazdów autonomicznych? Przykładowe obszary problemowe to: społeczne (neokolonializm poprzez narzucanie wartości, którymi ma kierować się AI), infrastrukturalne (niedostosowanie infrastruktury do potrzeb pojazdów autonomicznych), etyczne i prawne (kwestia podmiotowości AI, kwestia nadzoru nad pojazdem oraz odpowiedzialność cywilna i karna za ewentualne szkody). Jak widać na podstawie tego krótkiego wyliczenia, powstanie samochodów autonomicznych spowodowało wiele kontrowersji, których wyliczenie mogłoby stanowić przedmiot odrębnej monografii. Z tego względu niniejszy artykuł ogranicza się do problemów etycznych (wartości, podmiotowość i problem wyjaśnialności) oraz prawnych (odpowiedzialność cywilna i karna za ewentualne szkody).

W podrozdziale 2. omówię kwestię wartości, ich zależności kulturowej oraz społecznych oczekiwań co do technologii autonomicznych pojazdów. W części trzeciej tekstu przedstawię obecnie dominujący paradygmat tworzenia AI oraz problem wyjaśnialności – kluczowy dla zrozumienia działania AI, od którego to zrozumienia zależy określenie odpowiedzialności za ewentualne szkody. W podrozdziale 4. omówię mój pogląd na temat autonomii i moralnego statusu AI. Natomiast w piątym przedstawię ocenę kwestii odpowiedzialności i propozycję częściowego rozwiązania w postaci „kontraktu AI” oraz niepaństwowego funduszu ubezpieczeniowego, mającego pokryć przynajmniej część wyrządzonych szkód.

2. Kontekst kulturowy

Jedną z kluczowych kwestii związanych z wdrożeniem pojazdów autonomicznych stanowią kulturowo uwarunkowane różnice w postrzeganiu samochodów. W niektórych krajach są one tylko narzędziem mającym przewieźć człowieka z jednego miejsca w drugie, natomiast w innych samochód jest symbolem

społecznego statusu³. Ponadto, obserwujemy znaczące różnice w stylu jazdy i sposobie zachowania się kierowców w różnych rejonach. Zakładamy, że AI powinna prowadzić pojazdy bezpiecznie, lecz pojęcie bezpieczeństwa i normy zachowania będą się różnić w zależności od obszaru. Rodzi to pytanie: czy AI kierująca pojazdem powinna uwzględniać te różnice i zachowywać się inaczej podczas jazdy w Polsce, a inaczej w USA?

Wspomniane różnice dodatkowo utrudniają określenie docelowego sposobu funkcjonowania autonomicznych pojazdów oraz roli, jaką powinny odgrywać. Nie jest oczywiste, do jakiego stopnia powinny być one autonomiczne – czy AI powinno być ograniczone do wspierania kierowcy, a może docelowym etapem jest zastąpienie ludzkich kierowców? Wydaje się, że druga opcja jest ideowo dominującym trendem, jednak w praktyce – ze względu na obecny stopień rozwoju technologii oraz rozwiązania legislacyjne – te urządzenia traktowane są jako systemy wspomagające kierowcę. Powoduje to szereg dodatkowych komplikacji, m.in. w przypadku szkody wyrządzonej przez samochód sterowany przez algorytmy (w trybie tzw. samodzielnej jazdy, gdzie kierowca nie musi trzymać rąk na kierownicy) to kierujący pojazdem człowiek ponosi odpowiedzialność nie tylko cywilną, ale także karną. Wątek odpowiedzialności zostanie rozwinięty w piątej części artykułu.

Ważnym zagadnieniem w kontekście kulturowych uwarunkowań są również oczekiwania społeczne wobec technologii. Na platformie nazwanej „Moral Machine” przeprowadzono *online* badanie dotyczące oczekiwań co do decyzji podejmowanych przez pojazdy autonomiczne. Kwestionariusz został przygotowany w 10 językach. Zebrano 40 milionów odpowiedzi od osób z 233 krajów. Okazało się, że występują istotne różnice w oczekiwaniach pomiędzy osobami z różnych rejonów i kultur⁴. Jednak uwzględnienie w zbiorach danych, na podstawie których AI podejmuje decyzje, wszystkich kulturowo uwarunkowanych preferencji jest dużym wyzwaniem technicznym i moralnym. Sprawę komplikuje kwestia, czy AI, nawet tylko sterująca pojazdem autonomicznym, może stać się narzędziem neokolonializmu narzucającym wartości zachodniej kultury, jeśli nie uwzględnia preferencji mieszkańców krajów z innych kręgów kulturowych. Wydaje się, że społeczna odpowiedzialność twórców rozwiązań technicznych wymaga wzięcia pod uwagę takich czynników⁵.

³ K. Arcimowicz, *Znaczenia symboliczne oraz inne funkcje samochodu we współczesnej rodzinie*, „Studia z Teorii Wychowania” 2019, vol. 3 (28), s. 111–128, <https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-5fe45c8f-32fd-492f-8220-f86801513402> [dostęp: 08.05.2024].

⁴ E. Awad et al., *The Moral Machine Experiment*, „Nature” 2018, vol. 563, s. 59–64, <https://www.doi.org/10.1038/s41586-018-0637-6>.

⁵ S. Krumdieck, *Tackling ‘Wicked’ Problems Calls for Engineers with Social Responsibility*, „Nature” 2024, vol. 629, s. 994–995, <https://doi.org/10.1038/d41586-024-01519-1>.

Jakkolwiek jesteśmy w stanie zrozumieć, a nawet wybaczyć, ludzkie słabość i zawodność, to jednak spodziewamy się, że roboty będą doskonałe i niezawodne – wszak maszyny się nie męczą, nie mają złego humoru, nigdy nie są niewyspane, ani się nie rozpraszają. Mimo to dobrze wiemy, że w rzeczywistości zarówno sprzęt, jak i oprogramowanie są zawodne. Ta kwestia ma szczególne znaczenie w przypadku autonomicznych pojazdów. Ze względu na takie oczekiwania użytkownicy wykazują wysoki poziom zaufania⁶ do technologii, mimo że nie rozumieją jej działania. Zaufanie przejawia się w skłonności do zawierzenia algorytmom swojego zdrowia i życia m.in. poprzez świadome niesprawowanie nadzoru nad jadącym pojazdem⁷. Szczególnie szokującym przykładem takiego działania było użycie pomarańczy, aby „oszukać” znajdujący się w kierownicy czujnik mający zmusić kierowcę do trzymania na niej rąk, nawet gdy funkcja *Autopilot* jest aktywna⁸. Sprawę pogarsza fakt chwalenia się takimi działaniami w mediach społecznościowych, co dodatkowo zachęca do poszukiwania sposobów obejścia zabezpieczeń⁹. Te z kolei okazują się banalnie łatwe do pokonania, co zostało wielokrotnie wykazane, a ponadto doprowadziło do śmierci co najmniej dwóch osób¹⁰.

Inną istotną społecznie kwestią związaną z autonomicznymi samochodami jest możliwość utraty pracy przez zawodowych kierowców. Na obecnym etapie rozwoju techniki problem wydaje się niewielki, ale w przyszłości może przybrać na sile. Współcześnie istniejące pojazdy nie są prawdziwie autonomiczne, a ich dopuszczenie do użytku na publicznych drogach wzbudza liczne kontrowersje¹¹. Jednak nawet w swojej obecnej formie stają się częściowo samodzielnym uczestnikiem ruchu drogowego. Oznacza to, że taki samochód zaczyna być czymś więcej niż tylko

⁶ E.R. Teoh, *What's in a Name? Driver's Perceptions of the Use of Five SAE Level 2 Driving Automation Systems*, „Journal of Safety Research” 2020, vol. 72, s. 145–151, <https://doi.org/10.1016/j.jsr.2019.11.005>.

⁷ S. Nordhoff et al., *(Mis-)use of Standard Autopilot and Full Self-Driving (FSD) Beta: Results from Interviews with Users of Tesla's FSD Beta*, „Frontiers in Psychology” 2023, vol. 14, s. 1–20, <https://doi.org/10.3389/fpsyg.2023.1101520>.

⁸ Ph. Tracy, *This Guy Used an Orange to Hack His Tesla into a Self-Driving Car*, Daily Dot, 22.05.2021, <https://www.dailydot.com/debug/tesla-orange-hack/> [dostęp: 08.05.2024].

⁹ A. Linja, T.I. Mamun, Sh.T. Mueller, *When Self-Driving Fails: Evaluating Social Media Posts Regarding Problems and Misconceptions about Tesla's FSD Mode*, „Multimodal Technologies and Interaction” 2022, vol. 6 (10), 86, <https://doi.org/10.3390/mti6100086>.

¹⁰ S. Liberatore, *Tesla's Autopilot Can Be 'Easily Tricked' into Operating without a Driver, Says New Report Released after Two Men Die in a Fiery Model Y Crash with No One at the Wheel*, Daily Mail, 23.04.2021, <https://www.dailymail.co.uk/sciencetech/article-9503763/Tesla-car-tricked-operating-without-driver.html> [dostęp: 29.05.2024].

¹¹ N. Solis, R. Uranga, K. Garcia, *Waymo is Cleared to Launch Robotaxi Service in Los Angeles, Sparking New Protests*, Los Angeles Times, 01.03.2024, <https://www.latimes.com/california/story/2024-03-01/waymo-cleared-to-launch-robotaxi-service-in-los-angeles> [dostęp: 12.05.2024].

narzędziem. Nie tylko służy do konkretnego celu, jakim jest transport osób i rzeczy, lecz także może niezależnie podejmować decyzje, w wyniku których ktoś może ucierpieć. Z tej przyczyny pojazdy autonomiczne traktuję jako ambiwalentne, a nie neutralne pod względem moralnym. Wynika to z charakteru tej technologii oraz z nasycenia techniki wartościami¹² – niezależnie od możliwości wykorzystania jej w dobrym lub złym celu albo od nieprawidłowości jej używania. Kwestia możliwości uznania pojazdu autonomicznego za „uczestnika ruchu drogowego” w świetle polskiego prawa zasługuje na odrębne opracowanie.

3. AI jako „czarna skrzynka”

Obecnie dominującym paradygmatem w AI jest tworzenie modeli statystycznych na podstawie ogromnych ilości danych w procesie zwanym uczeniem maszynowym (ang. *machine learning*, ML). Oznacza to, że AI jest tak dobra, jak dobre są zbiory danych, na podstawie których była „trenowana”¹³. Od decyzji twórców tych systemów zależy, czy w takim zbiorze umieszczą więcej tzw. warunków brzegowych, czy skupią się na codziennych sytuacjach¹⁴. Warunki brzegowe to np. słynny dylemat wagonika oraz sytuacje związane z wypadkami¹⁵. Trenowanie modeli wymaga dużych ilości danych, dużej mocy obliczeniowych oraz ogromnej ilości energii, co także jest problemem z zakresu etyki środowiska naturalnego. Im więcej danych zostanie zawartych w zbiorze, tym większy staje się model ML i tym większej mocy obliczeniowej oraz pamięci potrzeba, aby mógł wykonywać swoje funkcje. Ponadto, duże modele działają wolniej, co jest szczególnie istotne w tych zastosowaniach, gdzie wymaga się natychmiastowej reakcji, np. w pojazdach autonomicznych. Konieczny jest zatem pewien kompromis między skutecznością takiego modelu (zależną od rodzaju i ilości danych) oraz szybkością jego działania (zależną od dostępnej pamięci i mocy obliczeniowej).

Ze względu na probabilistyczny charakter tych modeli ich działanie bywa nieprzewidywalne, co może się przejawiać jako niespodziewane hamowanie czy nieza-

¹² R. Lizut, *Technika a wartości. Spór o aksjologiczną neutralność artefaktów*, Wydawnictwo Akademicon, Lublin 2014.

¹³ E. Ferrara, *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*, „Sci” 2024, vol. 6 (1), <https://doi.org/10.3390/sci6010003>.

¹⁴ J. Basl, J. Behrends, *Why Everyone Has It Wrong about the Ethics of Autonomous Vehicles*, „The Bridge” 2019, vol. 49, issue 4, s. 42–47, <https://www.nae.edu/221212/Why-Everyone-Has-It-Wrong-about-the-Ethics-of-Autonomous-Vehicles> [dostęp: 15.04.2024].

¹⁵ J. Himmelreich, *Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations*, „Ethical Theory Moral Practice” 2018, vol. 21, s. 669–684, <https://doi.org/10.1007/s10677-018-9896-4>.

trzymanie się na znaku „stop”¹⁶. Nikt nie wie, jaką decyzję w konkretnych okolicznościach podejmie AI, co może istotnie wpływać na bezpieczeństwo – zachowanie pojazdu nie powinno być zaskakujące dla kierowcy i innych uczestników ruchu drogowego. Ta nieprzejrzystość wewnętrznego działania bywa określana jako „problem czarnej skrzynki” (ang. *black box problem*). W skrócie można to opisać następująco: znamy dane wejściowe (konkretną sytuację na drodze i odczyty z czujników w pojeździe) oraz wyjściowe (konkretna decyzja dotycząca zachowania pojazdu podjęta przez AI), ale nikt nie wie, na jakiej podstawie model doszedł do takich, a nie innych decyzji. Jest to szczególnie widoczne w modelach opartych na tzw. głębokim uczeniu (ang. *deep learning*), które są wykorzystywane w pojazdach autonomicznych. W ich przypadku między warstwą przyjmującą dane wejściowe i warstwą wyjściową znajduje się wiele tzw. warstw ukrytych. Prześledzenie krok po kroku działań wykonywanych przez algorytm w każdej z warstw jest niewykonalne ze względu na dużą złożoność matematycznych formuł za nimi stojących – są to nieliniowe równania z setkami, a nawet tysiącami zmiennych. To zbyt wiele, aby człowiek mógł to objąć umysłem. Jednym z przykładów „problemu czarnej skrzynki” w praktyce jest kazus z 2022 r., gdzie kierowca Tesli włączył funkcję *Autopilot* i zajął się przeglądaniem zawartości telefonu. Mimo dobrej widoczności samochód nie zatrzymał się w porę i śmiertelnie potrafił motocyklistę¹⁷. W tym przypadku to kierowca został przez sąd uznany za winnego, ale zdarzają się także sytuacje, gdy to producent pojazdu zgodził się wypłacić odszkodowanie¹⁸. Ze względu na dość wczesny etap rozwoju tej technologii pojazdy autonomiczne doświadczają wielu problemów¹⁹ i wciąż potrzebują nadzoru człowieka²⁰.

AI jest tworzona w „warunkach laboratoryjnych”, jednak używa się jej także poza laboratorium. Wczesny etap rozwoju i wielość problemów towarzyszących wdrażaniu technologii są jasnym sygnałem, że nie jest jeszcze gotowa do samodzielnego działania w warunkach „prawdziwego” świata. Budzi to zastrzeżenia co do realnego

¹⁶ G.A. Fowler, *Testing Tesla's Autopilot Recall, I Don't Feel Much Safer – and Neither Should You*, The Washington Post, 31.12.2023, <https://www.washingtonpost.com/technology/2023/12/31/tesla-autopilot-recall-test/> [dostęp: 12.05.2024].

¹⁷ B. Brownell, *Tesla Driver Charged with Killing Motorcyclist after Turning on Autopilot and Browsing His Phone*, Gizmodo, 24.04.2024, <https://gizmodo.com/tesla-motorcycle-crash-death-autopilot-washington-1851428850> [dostęp: 12.05.2024].

¹⁸ R. Felton, R. Elliott, *Tesla Agrees to Settle Lawsuit Over Autopilot's Involvement in 2018 Fatal Crash*, The Wall Street Journal, 08.04.2024, <https://www.wsj.com/business/autos/tesla-fatal-autopilot-crash-trial-33936b42> [dostęp: 12.05.2024].

¹⁹ R. Stumpf, *'Full Self-Driving' Teslas Keep Slamming Into Curbs*, Inside EVs, 14.04.2024, <https://insideevs.com/news/715913/tesla-fsd-trial-curb-hopping/> [dostęp: 15.04.2024].

²⁰ P. Berger, *Autonomous Delivery and Work Drones Will Still Need a Human Minder*, The Wall Street Journal, 07.11.2022, <https://www.wsj.com/articles/why-autonomous-vehicles-will-still-need-a-human-minder-11667833922> [dostęp: 15.05.2024].

poziomu bezpieczeństwa²¹. Dodatkową trudnością jest również niedopasowanie infrastruktury do potrzeb takich pojazdów – sieć ładowania samochodów elektrycznych nie jest jeszcze rozległa, a problemy z łącznością, objazdy i przeszkody mogą doprowadzić do niespodziewanego zatrzymania pojazdu²². Jest to widoczne np. w problemach trapiących autonomiczne taksówki, które nie zawsze radzą sobie z poruszaniem się po mieście, co skutkuje blokowaniem dróg oraz stłuczkami²³. Artykuł 65k p.r.d.²⁴ definiuje pojazd autonomiczny jako „[...] pojazd samochodowy, wyposażony w systemy sprawujące kontrolę nad ruchem tego pojazdu i umożliwiające jego ruch bez ingerencji kierującego, który w każdej chwili może przejąć kontrolę nad tym pojazdem”. Kluczowa dla tematu niniejszego artykułu jest kwestia osoby kierującej oraz możliwości sprawowania kontroli nad pojazdem. Dość oczywiste jest, że osoba w fotelu kierowcy odpowiada za sprawowanie nadzoru nad poruszającym się pojazdem, nawet jeśli ten bywa określany jako autonomiczny. Jednak istotna komplikacja pojawia się w przypadku autonomicznych taksówek, od użytkowników których oczekuje się zachowania, jakby byli tylko pasażerami. Taka sytuacja ma miejsce choćby w przypadku taksówek firmy Waymo, gdzie od użytkowników oczekuje się zajęcia miejsca na tylnej kanapie, a drzwi od strony kierowcy często pozostają zablokowane. Nie jest jasne, czy na gruncie polskiego prawa pasażer takiej taksówki mógłby zostać uznany przynajmniej za częściowo winnego kolizji z powodu niesprawowania nadzoru nad poruszającym się pojazdem, tylko dlatego, że w samochodzie znajduje się standardowa kierownica i pedały umożliwiające przejęcie nad nim kontroli. Obecnie nie ma przepisów jednoznacznie określających kwestię odpowiedzialności w takich przypadkach.

Z przedstawionej charakterystyki AI oraz kazusów wynikają dwa kluczowe problemy etyczne: 1) nie wiemy, jak AI podejmuje decyzje; 2) zwykle nie wiemy, czy decyzja podjęta przez AI faktycznie jest najlepsza w danej sytuacji. Pierwszy problem wynika z tego, że nawet najbardziej zaawansowane obecnie modele AI nie przedstawiają uzasadnień swoich działań, a my nie rozumiemy szczegółów mechanizmów za nimi stojących. Drugi natomiast wynika z braku wiedzy co do czynników, jakie AI bierze pod uwagę, a jakie pomija podczas podejmowania konkretnych

²¹ A. Momen et al., *Group Trust Dynamics during a Risky Driving Experience in a Tesla Model X*, „Frontiers in Psychology” 2023, vol. 14, s. 1–15, <https://doi.org/10.3389/fpsyg.2023.1129369>.

²² *Waymo Self-Driving Taxi Confused by Traffic Cones Flees Help*, BBC, 17.05.2021, <https://www.bbc.com/news/technology-57144755> [dostęp: 12.05.2024].

²³ M. Roeloffs, *Driverless Cars Face Setbacks in San Francisco – Here’s What to Know about the City’s Problematic Robotaxi Rollout*, Forbes, 21.08.2023, <https://www.forbes.com/sites/maryroeloffs/2023/08/21/driverless-cars-face-setbacks-in-san-francisco-heres-what-to-know-about-the-citys-problematic-robotaxi-rollout> [dostęp: 15.05.2024].

²⁴ Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 21 czerwca 2024 r. w sprawie ogłoszenia jednolitego tekstu ustawy – Prawo o ruchu drogowym (t.j. Dz.U. z 2024 r., poz. 1251).

decyzji. Rozwiązanie tych i innych problemów jest przedmiotem postulatu wyjaśnialnej AI (ang. *explainable AI*)²⁵, który zakłada takie projektowanie systemów, aby ich działanie było zrozumiałe dla człowieka²⁶. Oczywiście, dodatkową trudnością jest tu określenie, w jaki sposób ocenić, czy prezentowane wyjaśnienia i uzasadnienia faktycznie pozwalają użytkownikom tych systemów pojąć ich działanie²⁷.

Co istotne, warunku postawionego w drugiej części definicji z art. 65k p.r.d. nie spełniają pojazdy zupełnie pozbawione takich elementów, jak kierownica czy pedały, pozwalających człowiekowi na sterowanie. Przykładami takich samochodów są autonomiczne taksówki firmy Zoox²⁸, od listopada 2023 r. poruszające się po publicznych drogach Kalifornii, oraz pojazdy Origin tworzone przez General Motors, których wdrożenie obecnie zostało opóźnione ze względu na ryzyko związane z brakiem wystarczająco przejrzystych regulacji prawnych²⁹. W przypadku takich systemów dość oczywiste jest, że osoba znajdująca się w pojeździe nie ponosi moralnej odpowiedzialności za ewentualne szkody, ponieważ nie ma możliwości sprawowania nad nim kontroli. Jednak na gruncie polskiego prawodawstwa nie jest jasne, czy taka osoba może ponieść odpowiedzialność karną, ponieważ nierozstrzygnięty jest problem stopnia kontroli bądź wpływu, jaki może wyrzucić na pojazd osoba znajdująca się w nim. Jeśli samochód byłby wyposażony w hamulec awaryjny albo inny system umożliwiający zatrzymanie go przez pasażera, wówczas ten prawdopodobnie ponosiłby prawne konsekwencje braku użycia takiego systemu. Konieczne w tym przypadku jest rozstrzygnięcie, czy możliwość zatrzymania pojazdu jest tożsama ze sprawowaniem nad nim kontroli. Problematyczna natomiast zostaje kwestia, czy całość odpowiedzialności spoczywa na producencie (bądź importerze), czy też AI może zostać uznana za winną, przynajmniej częściowo. Kwestia odpowiedzialności samej AI jest szczególnie istotna, jeśli weźmiemy pod uwagę, że takie pojazdy posiadają techniczną możliwość poruszania się bez obecności człowieka w środku oraz przy minimalnym nadzorze zdalnym bądź zupełnie bez takiego nadzoru.

²⁵ F. Xu et al., *Explainable AI: a Brief Survey on History, Research Areas, Approaches and Challenges*, in: *Natural Language Processing and Chinese Computing. NLPCC 2019. Lecture Notes in Computer Science*, vol. 11839, eds. J. Tang et al., Springer, Cham 2019, s. 563–574, https://doi.org/10.1007/978-3-030-32236-6_51.

²⁶ A. Kuznetsov et al., *Explainable AI for Safe and Trustworthy Autonomous Driving: a Systematic Review*, arXiv, 08.02.2024, <https://doi.org/10.48550/arXiv.2402.10086>.

²⁷ R.R. Hoffman et al., *Metrics for Explainable AI: Challenges and Prospects*, arXiv, 01.02.2019, <https://doi.org/10.48550/arXiv.1812.04608>, <https://arxiv.org/pdf/1812.04608> [dostęp: 12.05.2024].

²⁸ *Zoox Purpose-Built Robotaxi is First in the World to Operate on Public Roads*, <https://www.datocms-assets.com/106048/1696542493-zoox-press-release-immediate-release.pdf> [dostęp: 15.05.2024].

²⁹ D. Shepardson, *GM Slams Brakes on Self-Driving Vehicle without Steering Wheel*, Reuters, 24.07.2024, <https://www.reuters.com/business/autos-transportation/gm-puts-self-driving-vehicle-without-steering-wheel-hold-2024-07-23/> [dostęp: 24.09.2024].

4. Implikacje pojęcia „autonomiczny”

Filozoficzne problemy z AI zaczynają się już w momencie sformułowania nazwy, która od samego początku jest wieloznaczna. „Sztuczna inteligencja” odnosi się zarówno do utworzonej w 1956 r. dyscypliny, jak i do głównego jej wytworu. Niestety ani dyscyplina, ani wytwór nie mają jednoznacznej i powszechnie akceptowanej definicji. Sam termin „inteligencja” ze względu na swoje konotacje (rozum i inteligencja są tradycyjnie uznawane za supranaturalne) nasuwa pytania o to, czy wykazywanie inteligentnego zachowania przez algorytmy i/lub sterowane przez nie urządzenia implikuje posiadanie umysłu, bądź czy tego wymaga. Podobnie jest z pojęciem „autonomia”, oznaczającym w kontekście AI pewien stopień niezależności i podmiotowości, które z kolei sugerują posiadanie jakiegoś rodzaju świadomości, co pod względem etycznym i prawnym wiąże się z objęciem ochroną, a być może także nałożeniem obowiązków³⁰. Określenie pojazdów sterowanych przez AI jako „autonomiczne” wskazywałoby, że mogą one działać bez nadzoru. Jednak nie jest to prawda, co widać po kazusach omówionych w poprzednich częściach tekstu.

Problemy terminologiczne wynikają również z faktu, że w opisach dotyczących AI oraz jej praktycznych realizacji, np. w postaci pojazdów autonomicznych, miesza się pojęcia odnoszące się do przyszłego, docelowego kształtu tej technologii oraz pojęcia opisujące jej obecny stan. Chyba najbardziej jaskrawym przykładem praktycznych konsekwencji tego zamieszania jest toczący się w USA pozew zbiorowy przeciwko firmie Tesla, która zapowiadała udostępnienie w swoich pojazdach opcji „w pełni autonomicznej jazdy”³¹, podczas gdy było wiadome, że ówczesne i obecne możliwości techniczne na to nie pozwalają. Optymistyczny marketing takich samochodów dodatkowo zachęca do nieostrożnych zachowań, które negatywnie wpływają na poziom bezpieczeństwa³². Nazwy wprowadzanych funkcji, takie jak: *Autopilot*, *Enhanced Autopilot* czy *Full Self Driving*, sugerują znacznie większe możliwości pojazdu, niż faktycznie on posiada³³. Z jednej strony podważa to deklarowaną przez producentów dbałość o bezpieczeństwo³⁴, a z dru-

³⁰ S. Schneider, *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*, przeł. J. Bednarek, Wydawnictwo Naukowe PWN, Warszawa 2021, s. 62–63.

³¹ J. Stempel, *Tesla Must Face Vehicle Owners' Lawsuit over Self-Driving Claims*, Reuters, 16.05.2024, <https://www.reuters.com/legal/tesla-must-face-vehicle-owners-lawsuit-over-self-driving-claims-2024-05-15/> [dostęp: 20.05.2024].

³² K. Poland et al., *Fatal Crash between a Car Operating with Automated Control Systems and a Tractor-Semitrailer Truck*, „Traffic Injury Prevention” 2018, vol. 19, s. 153–156, <https://doi.org/10.1080/15389588.2018.1532211>.

³³ E.R. Teoh, *What's in a Name?*...

³⁴ H. Wang et al., *A Survey on an Emerging Safety Challenge for Autonomous Vehicles: Safety of the Intended Functionality*, „Engineering” 2024, vol. 33, s. 17–34, <https://doi.org/10.1016/j.eng.2023.10.011>.

giej strony prowadzi do pytań o granice odpowiedzialności producenta urządzeń i ich użytkowników.

Artykuł 3 Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689³⁵ definiuje „system AI” jako „system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne”. Ponadto, regulacje dotyczące systemów AI zostały podzielone ze względu na szacowany poziom ryzyka, będący wypadkową prawdopodobieństwa wystąpienia szkody oraz potencjalnej dotkliwości tejże szkody. Tutaj chciałbym zauważyć, że zastosowanie AI w systemach sterujących pojazdem znacząco utrudnia ustalenie odpowiedzialnego za ewentualne szkody. Jest to spowodowane faktem, że tworzenie, modyfikacja i wdrażanie modeli AI angażują wiele podmiotów – od badaczy tworzących algorytmy stanowiące podstawę modeli AI, przez osoby tworzące (trenujące) modele AI, po zespoły wdrażające te modele w produktach, jak np. pojazdy autonomiczne³⁶. Zazwyczaj poszczególne etapy są realizowane przez podmioty należące do różnych grup, co dodatkowo komplikuje sytuację i może prowadzić do rozmywania odpowiedzialności. Kwestią problematyczną może się okazać wykazanie związku przyczynowego między działaniem twórcy a powstałym skutkiem³⁷. I należy zauważyć, że w praktyce może się to okazać niemożliwe. W przypadku pojazdów autonomicznych wykorzystujących wiele systemów opartych na AI, których elementy mogą pochodzić od różnych twórców, trudno jest określić, kto jest odpowiedzialny za konkretny fragment systemu, a więc kto ostatecznie może być uznany za jego twórcę. Ponadto, ze względu na skomplikowany charakter tej technologii wskazanie konkretnego czynnika bądź czynników prowadzących do powstania szkody może być niewykonalne („problem czarnej skrzynki”). Z tego powodu konieczne wydaje się opracowanie jednoznacznie określonego zakresu odpowiedzialności, ponieważ poleganie na kazuiście może w wielu przypadkach okazać się niewykonalne w praktyce. Ciekawym i wartym rozważenia kazusem są przedsiębiorstwa OEM (ang. *Original Equipment Manufacturer*), wytwarzające produkty, które następnie sprzedawane są przez inne przedsiębiorstwa pod ich własnymi markami. Kwestią do ustalenia pozostaje,

³⁵ Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r., https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=OJ:L_202401689 [dostęp: 24.09.2024].

³⁶ J.S. Borg, W. Sinnott-Armstrong, V. Conitzer, *Moral AI: and How We Get There*, Pelican Books, London 2024, s. 37–39.

³⁷ R. Rejmaniak, *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*, „Roczniki Nauk Prawnych” 2021, t. 31, nr 3, s. 104.

czy podział odpowiedzialności w takich okolicznościach może się istotnie różnić w przypadku, gdy dany produkt wykorzystuje AI, w porównaniu z przypadkiem, gdy produkt takich systemów nie wykorzystuje.

W obecnym stanie rozwoju technologii AI nie jest w pełni autonomiczna ani w kontekście etycznym, ani technicznym. Poziom niezależności od człowieka, mający znaczenie dla prawnej oceny AI, został określony na skali od 0 do 5, gdzie 5 oznacza pełną autonomię³⁸. Na chwilę obecną pełna autonomia jest technicznie nieosiągalna, a nawet niedopuszczalna prawnie, gdyż obowiązujące regulacje zakładają konieczność obecności człowieka w łańcuchu decyzyjnym. Choć w wielu przypadkach jako „autonomiczne” określa się systemy działające bez ingerencji człowieka, który zwykle pełni funkcję „bezpiecznika”, to nadal jest to tylko narzędzie stworzone przez ludzi, którzy decydują o jego rozwoju. Takie ujęcie nie umniejsza faktu, że jest to jedno z najbardziej zaawansowanych oraz wyrafinowanych narzędzi wytworzonych przez człowieka w celu zwiększenia możliwości i efektywności pracy. Jednak, jako narzędzie, AI nie może być moralnym podmiotem – nie jest ani dobra, ani zła. W takich kategoriach może być oceniane tylko wykorzystanie tej technologii przez ludzi. Algorytmy sterujące obecnie istniejącymi pojazdami autonomicznymi są przykładem słabej (wąskiej) AI – jest to określenie systemu wyspecjalizowanego do wykonywania jednego typu zadań. Podejmowane próby moralnej ewaluacji AI mogą się zatem odnosić tylko do tzw. silnej AI (AGI, ang. *Artificial General Intelligence*), która ma osiągnąć poziom „inteligencji” co najmniej równy ludziom, bądź do tzw. superinteligencji, która ma przekroczyć możliwości wszystkich ludzi. Takie koncepcje powodują wiele kontrowersji – jedni uważają, że AGI nie powstanie nigdy³⁹, a inni spodziewają się jej nadejścia już w 2029 r.⁴⁰. Co istotne, wąska AI nie jest w stanie uchwycić szerszego kontekstu, przez co ma trudności z podejmowaniem trafnych decyzji⁴¹. Takich problemów ma być pozbawiona AGI, która ma poprawnie wykonywać większość (albo wszystkie) rodzajów zadań na poziomie równym bądź lepszym od człowieka. Oczywiście, takie zadania obejmują także prowadzenie pojazdów. Niestety, AGI również nie doczekała się jednoznacznej definicji, a badacze z DeepMind zauważają, że dla zaistnienia AGI dany system nie musi

³⁸ T. Zalewski, *Definicja sztucznej inteligencji*, w: *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, C.H.Beck, Warszawa 2020, s. 11–13.

³⁹ J. Landgrebe, B. Smith, *Why Machines Will Never Rule The World. Artificial Intelligence without Fear*, Routledge, New York–London 2023.

⁴⁰ R. Kurzweil, *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii*, przeł. E. Chodkowska, A. Nowosielska, Kurhaus Publishing, Warszawa 2016.

⁴¹ J. McKendrick, A. Thurai, *AI Isn't Ready to Make Unsupervised Decisions*, Harvard Business Review, 15.09.2022, <https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions> [dostęp: 12.05.2024].

posiadać autonomii⁴². Zatem nawet powstanie takiej technologii niekoniecznie będzie oznaczało, że stanie się ona niezależnym od człowieka podmiotem, mogącym samodzielnie podejmować decyzje oraz ponosić moralne i prawne konsekwencje dokonywanych wyborów. Tutaj pozwolę sobie zakończyć rozważania, aby nie odbiegać zbyt od głównego tematu tekstu. Podkreślę, że najważniejszą konkluzją tej części jest, iż obecnie istniejące systemy AI nie są autonomiczne i nie mogą być traktowane jako podmiot w sensie moralnym ani prawnym.

5. Możliwe rozwiązania: „kontrakt AI”, prawa robotów

Z rozważań zawartych we wcześniejszych częściach artykułu wynika, że AI na obecnym poziomie rozwoju nie można przypisywać żadnego stopnia odpowiedzialności za podejmowane decyzje i działania wykonywane na ich podstawie, mimo nazywania jej „autonomiczną”. Ponadto, probabilistyczny charakter modeli stworzonych w dominującym obecnie paradygmacie nasuwa pytania o odpowiedzialność ich twórców za ewentualne szkody. Jest to istotne, biorąc pod uwagę fakt, że kierujący pojazdem określanym jako autonomiczny ponosi nieproporcjonalnie większe ryzyko niż producent takich urządzeń. Chodzi oczywiście o możliwość poniesienia odpowiedzialności karnej przez kierującego pojazdem, którego „autonomiczność” w żadnym stopniu nie umniejsza roli kierowcy w zdarzeniu drogowym. W obecnej formie AI nie może być podmiotem w świetle prawa, a producent pojazdu autonomicznego nie ponosi ryzyka większego niż producenci „zwykłych” samochodów. Praktycznie całe ryzyko spoczywa na kierującym. Kwestia ewentualnej odpowiedzialności przedsiębiorstwa i związanych z nim osób może być rozważana z wielu perspektyw. Aby nieco ograniczyć długość tekstu, podejmę się tylko wskazania możliwości uznania pojazdu autonomicznego za produkt niebezpieczny oraz odpowiedzialności osób zarządzających przedsiębiorstwem za nieuczciwą reklamę.

Skuteczność systemów AI stosowanych w pojazdach autonomicznych znacząco się różni w zależności od otoczenia pojazdu, ponieważ zależy od ilości i jakości danych, na których konkretny system był trenowany. Stąd jest niższa poza ścisłym centrum miast. Wynika to z faktu, że rzadziej uczęszczane drogi nie są tak dobrze „zmapowane” jak bardziej popularne trasy, co bezpośrednio przekłada się na ilość dostępnych danych. Wąska AI nie radzi sobie z sytuacjami, które nie pojawiły się w zbiorze, oraz ma problem ze zrozumieniem zmieniającego się kontekstu. W związku z tym poziom bezpieczeństwa zmienia się wraz ze zmianą otocze-

⁴² M.R. Morris et al., *Levels of AGI for Operationalizing Progress on the Path to AGI*, arXiv, 05.06.2024, <https://arxiv.org/pdf/2311.02462> [dostęp: 24.07.2024].

nia oraz panujących warunków. Na dwóch sąsiadujących ze sobą ulicach poziom skuteczności tego samego systemu może się znacząco różnić. Bez wprowadzenia „kontraktu AI” użytkownik może nie być świadomy, iż jego pojazd może przejawiać różne zachowania tylko dlatego, że dana droga częściej lub rzadziej pojawiała się w zbiorze danych, na których trenowano konkretną wersję AI. Prowadzi to do pytania: czy w związku z tym taki pojazd może zostać uznany za produkt niebezpieczny w świetle § 3 art. 449 k.c.⁴³? Zwłaszcza jeśli weźmie się pod uwagę fakt, że niektóre pojazdy nie stosują czujników będących standardem w podobnych konstrukcjach⁴⁴. Jeśli udzielimy odpowiedzi twierdzącej, to napotkamy kolejny problem: czy w przypadku autonomicznych samochodów ich oprogramowanie jest uznawane za integralną część, czy za coś odrębnego? Jest to istotne, ponieważ o ile wadliwy lub niedostatecznie precyzyjny czujnik może zostać uznany za produkt niebezpieczny, o tyle oprogramowanie jest dobrem intelektualnym, które nie posiada materialnej postaci, a zatem nie jest rzeczą ruchomą, o której mówi § 2 art. 449 k.c. W takiej interpretacji twórcy pojazdów autonomicznych mogliby przyjąć stanowisko, że są *de facto* producentami oprogramowania, a w związku z tym ich wytwór nie może zostać uznany za produkt niebezpieczny, przynajmniej na gruncie polskiego prawa. Świadczy to o konieczności wypracowania odrębnych przepisów dotyczących pojazdów autonomicznych. Nawet jeśli oprogramowanie mogłoby zostać uznane za produkt na podstawie wskazanych lub odrębnych przepisów, to w przypadku AI tworzonej w obecnie dominującym paradygmacie modeli statystycznych nie można dokonać rzetelnej oceny powodowanych zagrożeń ze względu na nieprzewidywalność jej działania. Jak widać, problem wyjaśnialności jest nie tylko kwestią techniczną, ale także prawną i moralną.

Odpowiedzialność osób związanych z przedsiębiorstwem regulują m.in. przepisy kodeksu spółek handlowych oraz kodeksu karnego. Jednak większość przepisów odnosi się do przestępstw przeciwko mieniu, w wyniku których spółka bądź inny podmiot ponosi szkodę majątkową. Jakkolwiek można by się doszukiwać znamion nadużycia zaufania (art. 296 k.k.⁴⁵) w fałszywym marketingu, tj. sprzedawaniu pojazdu autonomicznego jako posiadającego większe niż faktyczne możliwości i/ lub wykorzystywaniu w reklamie opisu nie obecnie posiadanych możliwości i funk-

⁴³ Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 21 czerwca 2024 r. w sprawie ogłoszenia jednolitego tekstu ustawy – Kodeks cywilny (t.j. Dz.U. z 2024 r., poz. 1061).

⁴⁴ Pojazdy marki Tesla nie posiadają Lidaru, który jest przez innych producentów uważany za krytyczną część zespołu czujników, w jakie wyposażony jest pojazd, ponieważ pozwala na tworzenie trójwymiarowej mapy otoczenia. Por. M. Watts et al., *Lidar on a Chip Puts Self-Driving Cars in the Fast Lane*, IEEE Spectrum, 20.08.2023, <https://spectrum.ieee.org/lidar-on-a-chip> [dostęp: 31.07.2024].

⁴⁵ Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 7 grudnia 2023 r. w sprawie ogłoszenia jednolitego tekstu ustawy – Kodeks karny (t.j. Dz.U. z 2024 r., poz. 17).

cji, lecz tych, których pojawienie się jest spodziewane w bliżej nieokreślonej przyszłości. Zwłaszcza gdy w kampanii marketingowej udzielają się osoby decyzyjne, czym narażają spółkę na szkodę majątkową z tytułu odszkodowań i zwrotu środków na rzecz klientów, co miało miejsce w przypadku pozwu przeciw firmie Tesla, gdzie za fałszywą reklamę odpowiadał m.in. jej prezes Elon Musk⁴⁶. Jednak w przypadku wyrządzenia szkody przez pojazd autonomiczny trudno byłoby wykazać odpowiedzialność konkretnych osób zarządzających przedsiębiorstwem, zwłaszcza w obecnej sytuacji, gdzie mamy do czynienia z technologią na dość wczesnym etapie rozwoju, której działania nie sposób przewidzieć. Tutaj mógłby mieć zastosowanie § 3 art. 483 k.s.h.⁴⁷, wedle którego działanie (tu: wytworzenie i wprowadzenie na rynek pojazdu autonomicznego) odbywa się w granicach uzasadnionego ryzyka gospodarczego, w tym informacji, które powinny być uwzględnione przy dokonywaniu oceny. Biorąc pod uwagę charakter tej technologii, można argumentować, że ewentualne szkody przez nią wyrządzone nie są celowym działaniem, lecz pewnego rodzaju skutkiem ubocznym jej rozwoju. Nie zmienia to jednak faktu, że brakuje przepisów określających udział przedsiębiorstwa w naprawie szkód.

Jedną z podstawowych kwestii związaną z ponoszeniem odpowiedzialności jest ocena ryzyka. Ze względu na sposób funkcjonowania technologii autonomicznych pojazdów dokonanie takiej oceny przez ich użytkownika jest w praktyce niewykonalne. Dane dotyczące (nie)zawodności takich samochodów nie są publicznie dostępne. W związku z tym osoby je kupujące stają się testerami technologii mogącej powodować niespodziewane zachowania pojazdu na drodze, co może stwarzać niebezpieczeństwo zarówno dla kierującego, jak i dla innych uczestników ruchu. Podstawowym czynnikiem ograniczania takiego ryzyka jest zrozumienie działania technologii oraz określenie zakresu odpowiedzialności w poszczególnych przypadkach. Stąd moją propozycją jest wprowadzenie „kontraktu AI”, który w realistyczny sposób przedstawiałby możliwości i ograniczenia technologii na obecnym etapie rozwoju oraz podział odpowiedzialności za ewentualne szkody między kierującym (lub użytkownikiem pojazdu) a producentem. Byłby to także element budowania społecznej świadomości poprzez odpowiednie działania edukacyjne, które mogłyby się przyczynić do zmniejszenia częstotliwości doświadczanych problemów oraz zwiększenia bezpieczeństwa na drogach⁴⁸. Odpowiedzialne wykorzystanie AI wymaga pewnego rodzaju umowy pomiędzy jej twórcami a użytkownikami.

Komplementarnym rozwiązaniem byłoby utworzenie niepaństwowego funduszu ubezpieczeniowego, w działaniu podobnego do Ubezpieczeniowego Funduszu

⁴⁶ J. Stempel, *Tesla Must Face...*

⁴⁷ Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 7 grudnia 2023 r. w sprawie ogłoszenia jednolitego tekstu ustawy – Kodeks spółek handlowych (t.j. Dz.U. z 2024 r., poz. 18).

⁴⁸ A. Linja, T.I. Mamun, Sh.T. Mueller, *When Self-Driving Fails...*

Gwarancyjnego. Jego celem byłoby pokrycie części szkód niezależnie od oceny stopnia odpowiedzialności producenta i użytkownika pojazdu. Byłby to rodzaj dodatkowego ubezpieczenia dla autonomicznych pojazdów, które docelowo mają się poruszać bez kierowcy. Państwo nie powinno, moim zdaniem, uczestniczyć w kompensacji szkód, ponieważ wprowadzenie pojazdów autonomicznych nie jest obszarem krytycznie ważnym dla funkcjonowania społeczeństwa. Publiczne środki mogą wspierać rozwój infrastruktury, co byłoby usprawiedliwione odnoszeniem społecznych korzyści, ponieważ mogłyby jej używać nie tylko pojazdy autonomiczne. Natomiast ewentualne szkody nie powinny być kompensowane z budżetu państwa, gdyż zyskałyby na tym tylko niektóre przedsiębiorstwa, a ponadto wsparcie byłoby nieproporcjonalnie większe dla firm mających największe udziały w rynku. Poza tym mogłoby to osłabiać inicjatywę dla firm, aby dążyć do minimalizacji ewentualnych szkód. Oczywiście, nawet niepaństwowy fundusz nie może być jedynym rozwiązaniem – potrzebne są działania kompleksowe obejmujące także legislację. Nie można doprowadzić do sytuacji, gdzie podpisanie kontraktu oraz opłacenie dodatkowego ubezpieczenia zdejmą z producentów i użytkowników całą odpowiedzialność za ewentualne szkody. W takim przypadku powstałoby ryzyko, że twórcy rozwiązań technicznych potraktowaliby to jako koszt prowadzenia działalności i nie mieliby odpowiedniej zachęty do dbania o podnoszenie poziomu bezpieczeństwa.

Przyjęte przez UE przepisy regulujące zastosowanie sztucznej inteligencji – AI Act⁴⁹ – w obecnej formie nie regulują wykorzystania autonomicznych pojazdów. Algorytmy za nimi stojące można zakwalifikować jako „systemy wysokiego ryzyka” wedle nomenklatury wprowadzonej w dokumencie, co ma nakładać na ich producentów obowiązek oceny i ograniczenia ryzyka oraz prowadzenia rejestrów zdarzeń. Co istotne, Akt zakłada realizację postulatu wyjaśnialności AI oraz konieczność jej kontroli przez człowieka⁵⁰. Jest to dość ogólne określenie obowiązków. Szczegółowe zasady zostaną opracowane w regulacjach sektorowych.

Kwestia, czy robota lub samochód faktycznie można by uznać za winnego lub współwinnego wyrządzonej szkody, gdy technologia osiągnie odpowiedni poziom rozwoju, wymaga dalszych badań. Aby to było możliwe, taki robot musiałby posiadać indywidualną osobowość moralną, a może także prawną. Wynika to z faktu, że wina i odpowiedzialność wiążą się ze zdolnością do podejmowania decyzji, oceny

⁴⁹ T. Madiega, *Artificial Intelligence Act*, EU Legislation in Progress, 03.2024, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf) [dostęp: 30.05.2024]. Pełny tekst rozporządzenia: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf [dostęp: 30.05.2024].

⁵⁰ *Akt ws. sztucznej inteligencji: pierwsze przepisy regulujące sztuczną inteligencję*, Parlament Europejski, 04.04.2024, <https://www.europarl.europa.eu/topics/pl/article/20230601STO93804/akt-ws-sztucznej-inteligencji-pierwsze-przepisy-regulujace-ai> [dostęp: 15.04.2024].

tych decyzji, świadomością skutków i ponoszenia konsekwencji. Natomiast konsekwencje mogą być zarówno negatywne, jak i pozytywne (nagrody i kary). Jeśli uznalibyśmy, że robot może odpowiadać za wyrządzone szkody, to powinniśmy również uznać, że powinien otrzymywać wynagrodzenie za wykonaną pracę. Co ciekawe, robot nie musi posiadać świadomości na takim samym poziomie, co ludzie. Brak świadomości nie wyklucza braku godności. Jeśli okazałoby się, że świadomość robota będzie na innym poziomie niż ludzka, to prawdopodobnie oznaczałoby to konieczność utworzenia odrębnego prawa dla robotów, podobnie jak teraz postuluje się ochronę praw zwierząt⁵¹. Takie rozwiązanie byłoby też bardziej korzystne, jako że minimalizuje ryzyko „erozji” praw dotyczących ludzi.

6. Podsumowanie

Obecnie pojazdy autonomiczne wciąż znajdują się w dość wczesnej fazie rozwoju. Nie jest jasne, czy i kiedy faktycznie będą w stanie działać w pełni samodzielnie. Problemy z nimi związane wskazują, że konieczne są badania inter- i transdyscyplinarne, zanim takie wytwory będą się nadawały do użytku nie tylko pod względem technicznym, ale także społecznym.

Rozstrzygnięcia prawne odnoszące się do sposobów zastosowania AI w takich pojazdach wydają się koniecznością. Obecnie wprowadzany Akt w sprawie sztucznej inteligencji tylko w ogólnym stopniu określa obowiązki dotyczące producentów takich rozwiązań. Na szczegółowe rozstrzygnięcia trzeba poczekać do czasu opracowania regulacji sektorowych. Zaproponowane w niniejszym artykule rozwiązania należy uznać za komplementarne. Wprowadzenie „kontraktu AI” nie może zastąpić prawnych regulacji, podobnie jak dodatkowe ubezpieczenie nie może być jedynym środkiem mającym minimalizować negatywne skutki wdrażania inteligentnych technologii.

Kwestią wymagającą dalszych badań pozostaje natomiast ujęcie AI jako „autonomicznej” oraz możliwość ponoszenia konsekwencji podjętych decyzji. Takie określenie implikuje posiadanie podmiotowości, a przypisanie winy bądź współwiny wiąże się z posiadaniem zdolności do podejmowania decyzji, świadomości skutków podjętych działań oraz ponoszeniem konsekwencji (zarówno pozytywnych, jak i negatywnych). Nie jest oczywiste, czy AI w ogóle może posiadać takie cechy.

Przedstawione w tekście analizy dotyczące odpowiedzialności użytkowników i twórców pojazdów autonomicznych z pewnością nie są wyczerpujące. Kwestia

⁵¹ J. Gellers, *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*, Routledge, London 2022.

możliwości uznania takiego pojazdu za produkt niebezpieczny oraz odpowiedzialność osób zarządzających przedsiębiorstwem na pewno wymagają dalszych rozważań. Zwłaszcza ze względu na konieczność wypracowania szczegółowych przepisów dotyczących autonomicznych samochodów. Jak widać, problem wyjaśnialności ma wymiar nie tylko techniczny, ale także etyczny i prawny.

Bibliografia

- Akt ws. sztucznej inteligencji: pierwsze przepisy regulujące sztuczną inteligencję*, Parlament Europejski, 04.04.2024, <https://www.europarl.europa.eu/topics/pl/article/20230601STO93804/akt-ws-sztucznej-inteligencji-pierwsze-przepisy-regulujace-ai> [dostęp: 15.04.2024].
- Arcimowicz K., *Znaczenia symboliczne oraz inne funkcje samochodu we współczesnej rodzinie*, „Studia z Teorii Wychowania” 2019, vol. 3 (28), s. 111–128, <https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-5fe45c8f-32fd-492f-8220-f86801513402> [dostęp: 08.05.2024].
- Awad E. et al., *The Moral Machine Experiment*, „Nature” 2018, vol. 563, s. 59–64, <https://doi.org/10.1038/s41586-018-0637-6>.
- Basl J., Behrends J., *Why Everyone Has It Wrong about the Ethics of Autonomous Vehicles*, „The Bridge” 2019, vol. 49, issue 4, s. 42–47, <https://www.nae.edu/221212/Why-Everyone-Has-It-Wrong-about-the-Ethics-of-Autonomous-Vehicles> [dostęp: 15.04.2024].
- Berger P., *Autonomous Delivery and Work Drones Will Still Need a Human Minder*, The Wall Street Journal, 07.11.2022, <https://www.wsj.com/articles/why-autonomous-vehicles-will-still-need-a-human-minder-11667833922> [dostęp: 15.05.2024].
- Borg J.S., Sinnott-Armstrong W., Conitzer V., *Moral AI: and How We Get There*, Pelican Books, London 2024.
- Brownell B., *Tesla Driver Charged with Killing Motorcyclist after Turning on Autopilot and Browsing His Phone*, Gizmodo, 24.04.2024, <https://gizmodo.com/tesla-motorcycle-crash-death-autopilot-washington-1851428850> [dostęp: 12.05.2024].
- Campbell C., *AI by Design: A Plan for Living with Artificial Intelligence*, CRC Press, New York 2022.
- Felton R., Elliott R., *Tesla Agrees to Settle Lawsuit Over Autopilot’s Involvement in 2018 Fatal Crash*, The Wall Street Journal, 08.04.2024, <https://www.wsj.com/business/autos/tesla-fatal-autopilot-crash-trial-33936b42> [dostęp: 12.05.2024].
- Ferrara E., *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*, „Sci” 2024, vol. 6 (1), <https://doi.org/10.3390/sci6010003>.

- Fowler G.A., *Testing Tesla's Autopilot Recall, I Don't Feel Much Safer – and Neither Should You*, The Washington Post, 31.12.2023, <https://www.washingtonpost.com/technology/2023/12/31/tesla-autopilot-recall-test/> [dostęp: 12.05.2024].
- Gellers J., *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*, Routledge, London 2022.
- Himmelreich J., *Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations*, "Ethical Theory Moral Practice" 2018, vol. 21, s. 669–684, <https://doi.org/10.1007/s10677-018-9896-4>.
- Hoffman R.R. et al., *Metrics for Explainable AI: Challenges and Prospects*, arXiv, 01.02.2019, <https://doi.org/10.48550/arXiv.1812.04608>, <https://arxiv.org/pdf/1812.04608> [dostęp: 12.05.2024].
- Krumdieck S., *Tackling 'Wicked' Problems Calls for Engineers with Social Responsibility*, "Nature" 2024, vol. 629, s. 994–995, <https://doi.org/10.1038/d41586-024-01519-1>.
- Kurzweil R., *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii*, przeł. E. Chodkowska, A. Nowosielska, Kurhaus Publishing, Warszawa 2016.
- Kuznietsov A. et al., *Explainable AI for Safe and Trustworthy Autonomous Driving: a Systematic Review*, arXiv, 08.02.2024, <https://doi.org/10.48550/arXiv.2402.10086>.
- Landgrebe J., Smith B., *Why Machines Will Never Rule The World. Artificial Intelligence without Fear*, Routledge, New York–London 2023.
- Liberatore S., *Tesla's Autopilot Can Be 'Easily Tricked' into Operating without a Driver, Says New Report Released after Two Men Die in a Fiery Model Y Crash with No One at the Wheel*, Daily Mail, 23.04.2021, <https://www.dailymail.co.uk/sciencetech/article-9503763/Tesla-car-tricked-operating-without-driver.html> [dostęp: 29.05.2024].
- Linja A., Mamun T.I., Mueller Sh.T., *When Self-Driving Fails: Evaluating Social Media Posts Regarding Problems and Misconceptions about Tesla's FSD Mode*, "Multimodal Technologies and Interaction" 2022, vol. 6 (10), 86, <https://doi.org/10.3390/mti6100086>.
- Lizut R., *Technika a wartości. Spór o aksjologiczną neutralność artefaktów*, Wydawnictwo Academicon, Lublin 2014.
- Madiega T., *Artificial Intelligence Act*, EU Legislation in Progress, 03.2024, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf) [dostęp: 30.05.2024].
- McKendrick J., Thurai A., *AI Isn't Ready to Make Unsupervised Decisions*, Harvard Business Review, 15.09.2022, <https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions> [dostęp: 12.05.2024].
- Momen A. et al., *Group Trust Dynamics during a Risky Driving Experience in a Tesla Model X*, "Frontiers in Psychology" 2023, vol. 14, s. 1–15, <https://doi.org/10.3389/fpsyg.2023.1129369>.
- Morris M.R. et al., *Levels of AGI for Operationalizing Progress on the Path to AGI*, arXiv, 05.06.2024, <https://arxiv.org/pdf/2311.02462> [dostęp: 24.07.2024].

- Nordhoff S. et al., *(Mis-)use of Standard Autopilot and Full Self-Driving (FSD) Beta: Results from Interviews with Users of Tesla's FSD Beta*, "Frontiers in Psychology" 2023, vol. 14, s. 1–20, <https://doi.org/10.3389/fpsyg.2023.1101520>.
- Parlament Europejski, *Artificial Intelligence Act*, 19.04.2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf [dostęp: 30.05.2024].
- Poland K. et al., *Fatal Crash between a Car Operating with Automated Control Systems and a Tractor-Semitrailer Truck*, "Traffic Injury Prevention" 2018, vol. 19, s. 153–156, <https://doi.org/10.1080/15389588.2018.1532211>.
- Rejmaniak R., *Autonomiczność systemów sztucznej inteligencji jako wyzwanie dla prawa karnego*, „Roczniki Nauk Prawnych” 2021, t. 31, nr 3, s. 95–113.
- Roeloffs M., *Driverless Cars Face Setbacks in San Francisco – Here's What to Know about the City's Problematic Robotaxi Rollout*, Forbes, 21.08.2023, <https://www.forbes.com/sites/maryroeloffs/2023/08/21/driverless-cars-face-setbacks-in-san-francisco-heres-what-to-know-about-the-citys-problematic-robotaxi-rollout> [dostęp: 15.05.2024].
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r., https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=OJ:L_202401689 [dostęp: 24.09.2024].
- Savage N., *How Robots Can Learn to Follow a Moral Code*, Nature, 26.10.2023, <https://www.nature.com/articles/d41586-023-03258-1> [dostęp: 30.05.2024].
- Schneider S., *Świadome maszyny. Sztuczna inteligencja i projektowanie umysłów*, przeł. J. Bednarek, Wydawnictwo Naukowe PWN, Warszawa 2021.
- Shepardson D., *GM Slams Brakes on Self-Driving Vehicle without Steering Wheel*, Reuters, 24.07.2024, <https://www.reuters.com/business/autos-transportation/gm-puts-self-driving-vehicle-without-steering-wheel-hold-2024-07-23/> [dostęp: 24.09.2024].
- Solis N., Uranga R., Garcia K., *Waymo is Cleared to Launch Robotaxi Service in Los Angeles, Sparking New Protests*, Los Angeles Times, 01.03.2024, <https://www.latimes.com/california/story/2024-03-01/waymo-cleared-to-launch-robotaxi-service-in-los-angeles> [dostęp: 12.05.2024].
- Stempel J., *Tesla Must Face Vehicle Owners' Lawsuit over Self-Driving Claims*, Reuters, 16.05.2024, <https://www.reuters.com/legal/tesla-must-face-vehicle-owners-lawsuit-over-self-driving-claims-2024-05-15/> [dostęp: 20.05.2024].
- Stumpf R., *'Full Self-Driving' Teslas Keep Slamming Into Curbs*, Inside EVs, 14.04.2024, <https://insideevs.com/news/715913/tesla-fsd-trial-curb-hopping/> [dostęp: 15.04.2024].
- Teoh E.R., *What's in a Name? Driver's Perceptions of the Use of Five SAE Level 2 Driving Automation Systems*, "Journal of Safety Research" 2020, vol. 72, s. 145–151, <https://doi.org/10.1016/j.jsr.2019.11.005>.
- Tracy Ph., *This Guy Used an Orange to Hack His Tesla into a Self-Driving car*, Daily Dot, 22.05.2021, <https://www.dailydot.com/debug/tesla-orange-hack/> [dostęp: 08.05.2024].

- Wang H. et al., *A Survey on an Emerging Safety Challenge for Autonomous Vehicles: Safety of the Intended Functionality*, "Engineering" 2024, vol. 33, s. 17–34, <https://doi.org/10.1016/j.eng.2023.10.011>.
- Watts M. et al., *Lidar on a Chip Puts Self-Driving Cars in the Fast Lane*, IEEE Spectrum, 20.08.2023, <https://spectrum.ieee.org/lidar-on-a-chip> [dostęp: 31.07.2024].
- Waymo Self-Driving Taxi Confused by Traffic Cones Flees Help, BBC, 17.05.2021, <https://www.bbc.com/news/technology-57144755> [dostęp: 12.05.2024].
- Xu F. et al., *Explainable AI: a Brief Survey on History, Research Areas, Approaches and Challenges*, in: *Natural Language Processing and Chinese Computing. NLPCC 2019. Lecture Notes in Computer Science*, vol. 11839, eds. J. Tang et al., Springer, Cham 2019, s. 563–574, https://doi.org/10.1007/978-3-030-32236-6_51.
- Zalewski T., *Definicja sztucznej inteligencji*, w: *Prawo sztucznej inteligencji*, red. L. Lai, M. Świerczyński, C.H.Beck, Warszawa 2020, s. 1–14.
- Zoox Purpose-Built Robotaxi is First in the World to Operate on Public Roads, <https://www.datocms-assets.com/106048/1696542493-zoox-press-release-immediate-release.pdf> [dostęp: 15.05.2024].

mgr DAMIAN SZCZĘCH

e-mail: damian.szczech@kul.pl

Uczestnik Szkoły Doktorskiej KUL na kierunku: filozofia, gdzie pisze pracę związaną z etyką sztucznej inteligencji. Zajmuje się zagadnieniami łączącymi się z transhumanizmem oraz nowymi technologiami.

Participant of the KUL Doctoral School in the field of philosophy, working on a thesis related to the ethics of artificial intelligence. He deals with the issues concerned with transhumanism and new technologies.