

Bezpieczeństwo przede wszystkim



Zamiast zastanawiać się, w jakim zakresie powinniśmy pozwolić sztucznej inteligencji podejmować samodzielnie decyzje moralne, trzeba zmierzyć się z problemem: jak mamy projektować pojazdy autonomiczne, by nie dopuścić do stawiania ich przed dylematami moralnymi.

Wraz z upowszechnianiem się zastosowań sztucznej inteligencji w różnych dziedzinach życia będą padały kolejne mity związane z jej wczesną fazą rozwoju. Moim zdaniem, nie wytrzyma próby czasu mit o konieczności podejmowania przez systemy bazujące na AI decyzji nacechowanych etycznie. A nawet szerzej, nie obroni się też dzisiejsze wyobrażenie o podejmowaniu przez AI jakichkolwiek decyzji w sensie takim, jak je podejmuje człowiek.

Jednym z obszarów wymagających regulowania etycznych zachowań sztucznej inteligencji jest sfera autonomicznego transportu. W dyskusjach o samojezdnym pojazdach wciąż jeszcze pojawiają się wyobrażenia, że systemy sterowania samochodami będą zmuszone do samodzielnego podejmowania tragicznych w skutkach decyzji, kogo zabić: czy wjechać na przykład w grupę przechodzących przez pasy dzieci, czy też skrócić ostro na pobocze i uderzyć w rosnące przy drodze drzewo uśmiercając wszystkich pasażerów. Padają przy tym



Andrzej Gontarz

Z wykształcenia kulturoznawca, w pracy zawodowej związany z branżą ICT, m.in. jako pełnomocnik dyrektora ds. informatyzacji, dziennikarz i redaktor prasy specjalistycznej, analityk trendów sektorowych, twórca raportów.

jednocześnie pytania o zasady moralne, jakimi w takich sytuacjach powinny się kierować jeżdżące bez kierowcy maszyny.

To, moim zdaniem, z gruntu fałszywe i nieuprawnione podejście. Etyka i moralność są właściwe człowiekowi a nie maszynie. Ponieważ jednak wciąż mamy skłonności do niesłusznego traktowania sztucznej inteligencji w kategoriach bytu jednostkowego, to automatycznie przypisujemy jej też cechy osobowe. To prowadzi nas w konsekwencji do tak absurdalnych przekonań i postaw jak oczekiwanie od tzw. inteligentnych maszyn umiejętności moralnego osądu zaistniałej sytuacji i podejmowania etycznych decyzji.

Nowa odsłona dylematu wagonika

Próba stworzenia swego rodzaju kręgosłupa moralnego dla pojazdów autonomicznych był realizowany w latach 2016–2020 przez MIT (Massachusetts Institute of Technology) projekt „Morale Machine”. Jego celem było zebranie opinii ludzi z całego świata na temat tego, jak w sytuacjach, w których nie można uniknąć wypadku, powinien zadziałać samochód bez kierowcy. Program ten miał charakter otwartego badania, w którym każdy chętny mógł przedstawić swoją propozycję decyzji dla wybranych scenariuszy drogowych z udziałem pojazdów autonomicznych. Uczestnicy badania mieli wskazać, kogo auto miałyby oszczędzić, a kogo poświęcić dla uratowania innych. Kto miałby przeżyć, a kto zginąć - młodszy czy starszy, pieszy czy pasażer, osoba zdrowa czy chora, jadący samochodem przestępca czy przechodzący na czerwonym świetle normalny obywatel? Takich dylematów do oceny było kilkanaście.

Projekt MIT stanowił odmianę czy też współczesną interpretację zaproponowanego w 1967 r. przez Philippę Foot eksperymentu myślowego znanego jako dylemat wagonika. Rozpędzony wagon z zepsutymi hamulcami pędzi w dół po torze, do którego szalenic przywiązał pięć osób. Wagonu nie można zatrzymać, ale można przestawić zwrotnicę i skierować go na tor, na którym znajduje się jeden człowiek. Co w tej sytuacji powinieneś zrobić? Tak zarysowana hipotetyczna sytuacja stała się pretekstem do wielu ważnych rozważań na temat problemów z obszaru etyki, moralności, systemów wartości i aksjomatów moralnych.

W obiegu jest też wiele innych, czasami prostszych, czasami bardziej skomplikowanych wersji tego typu dylematu, jak na przykład dylemat speleologów czy dylemat chirurga-transplantologa. Projekt „Morale machine” jest jedną z odmian tego rodzaju eksperymentów myślowych. Jego twórcy odwołują się też zresztą wprost do tych klasycznych przykładów. I właśnie jako eksperyment myślowy powinien być traktowany, a nie jako zbiór propozycji dla konstruktorów pojazdów autonomicznych. W rzeczywistości bowiem, tak jak wszystkie eksperymenty myślowe, unaocznia raczej złożoność, subtelną i niejednoznaczność rozważanego problemu niż pokazuje jego rozwiązanie.

Przynależność człowiekowi

W badaniu prowadzonym przez MIT nie chodziło w gruncie rzeczy o zbieranie propozycji do katalogu potencjalnie dopuszczalnych zachowań dla pojazdów autonomicznych. Chodziło raczej o poznanie zapatrywania się ludzi na niejednoznaczne moralnie sytuacje możliwe do zaistnienia w ruchu drogowym z udziałem samochodów bez kierowców. Chodziło o poznanie ludzkiej perspektywy aksjologicznej, uchwycenie granic akceptacji lub dezaprobaty dla możliwych rozstrzygnięć hipotetycznych sytuacji z udziałem ludzi i aut samojeżdżących. Wyniki nie dają podstaw do zaprojektowania uniwersalnych zasad moralnych dla maszyn autonomicznych, chociaż wynika z nich kilka generalnych trendów. Pozwalają natomiast zobaczyć w jak różnorodny sposób na kwestie wyboru w sytuacjach ekstremalnych zapatrują się różni ludzie na świecie.

Etyka i moralność jest bowiem przypisana ludziom, jest nieodłącznym aspektem ludzkiej obecności w świecie i ludzkiego funkcjonowania w środowisku społecznym. Tak samo jak związana z tym odpowiedzialność za podejmowane wybory i czyny. Człowiek nie może tej sfery swego istnienia scedować na maszyny, jak by niektórzy chcieli, nawet jeśli są inteligentne.

Jakie jest więc wyjście z sytuacji określanej umownie mianem dylematu wagonika w sferze transportu autonomicznego? Najprostsze z możliwych i w zasadzie jedyne dopuszczalne etycznie rozwiązanie, to konstruować auta bez kierowcy i sterujące nimi oprogramowanie w taki sposób, aby do powstawania sytuacji dwuznacznych moralnie w ogóle w ruchu ulicznym nie dochodziło. Czy to w stu procentach możliwe? Być może nie, ale tak jak w każdej innej dziedzinie inżynierii, tak i tu, w systemach sztucznej inteligencji montowanej w pojazdach, należy ustalić absolutne granice dopuszczalnego ryzyka i ich przestrzegać.

Minimalizacja ryzyka

Na samochody autonomiczne należy patrzeć jak na urządzenia techniczne a nie podmioty moralne. To jedyny sposób poradzenia sobie z etycznymi wyzwaniem związanymi z rozwojem tego obszaru zastosowań sztucznej inteligencji. Jedynym etycznie dopuszczalnym zachowaniem jest niedopuszczanie przez twórców pojazdów autonomicznych do sytuacji, które w jakikolwiek sposób wymagały podejmowania przez maszyny nacechowanych moralnie decyzji.

W jaki sposób można to zrobić? W pierwszym rzędzie najlepiej przez maksymalne wykorzystanie możliwości dostępnych rozwiązań technicznych. Inspiracje można czerpać chociażby z sektora lotnictwa czy obszaru automatyki przemysłowej na kolei, gdzie standardem jest

stosowanie rozwiązań redundantnych. Duplikowanie, czy wręcz multiplikowanie zastosowanych systemów i układów może w znaczący sposób zwiększać gwarancje osiągnięcia odpowiedniego poziomu bezpieczeństwa dla samochodów poruszających się samodzielnie po drogach. Liczy się tu przede wszystkim stabilność i przewidywalność działania wypuszczanych w przestrzeń publiczną maszyn.

Stosowana dzisiaj w sektorze biznesowym praktyka produkcji i wykorzystania samochodów autonomicznych zdaje się ten kierunek działania potwierdzać. – Podstawowym założeniem przyjmowanym przy konstruowaniu pojazdów autonomicznych jest to, że system ma za wszelką cenę minimalizować ryzyko. Musi przewidywać możliwe zachowania innych uczestników ruchu oraz unikać kolizji. Nie ma mowy o tym, by system sam decydował i wybierał w kogo ma uderzyć, jak to się często przedstawia w mediach. Po to na pojeździe montowanych jest tak wiele sensorów, żeby w każdych warunkach pojazd miał dostateczną ilość potrzebnych informacji do bezpiecznych reakcji, nawet przy złej pogodzie, czy w innych warunkach ograniczonej widoczności – tłumaczył w wywiadzie dla magazynu CRN (nr 10/2024) Łukasz Wójcik, CTO w gliwickiej firmie Bleeps. Przedsiębiorstwo jest producentem autonomicznego minibusa, który obecnie odbywa jazdy testowe w różnych polskich miastach.

Nie tylko technika

Życie ludzkie jest wartością bezwzględna, najwyższą i trzeba zrobić wszystko, aby je ochronić. To zadanie i obowiązek również twórców i użytkowników sztucznej inteligencji. Jeżeli nie da się tego osiągnąć przez dostępne rozwiązania techniczne, bo są na przykład jeszcze zbyt ułomne, żeby radzić sobie ze wszystkimi skomplikowanymi sytuacjami, to trzeba to zrobić w inny sposób, korzystając z innych dostępnych rozwiązań i możliwości. W przypadku transportu autonomicznego mogą to być na przykład rozwiązania prawne, infrastrukturalne, a nawet edukacyjne.

Na gruncie techniki trzeba stosować rozwiązania jednoznaczne, takie, które będą realizowały jednoznacznie określone zadania, powodowały uruchamianie jednoznacznych działań – w celu zapewnienia ludziom maksymalnego poziomu bezpieczeństwa. Można wprowadzić dodatkowe rozwiązania zabezpieczające – na przykład wydzielone pasy ruchu, zmiany w organizacji ruchu lub w sygnalizacji świetlnej, nowe regulacje prawne wymuszające odpowiednie zachowania pieszych względem samochodów bez kierowców czy normy społeczne zabraniające określonych zachowań, na przykład nagłego wchodzenia przed pojazd samosterujący.

Próby poszukiwania takich rozwiązań są już podejmowane. Pojawił się na przykład pomysł wprowadzenia dodatkowe-

go, białego światła na sygnalizatorach dającego pierwszeństwo przejazdu samochodom-automatom. Prowadzone są też prace nad ustandaryzowaniem i usystematyzowaniem na nowo organizacji ruchu drogowego, by dostosować jego zasady do nowych wyzwań związanych z obecnością aut autonomicznych. W szukaniu właściwych rozwiązań trzeba brać pod uwagę, z jednej strony, możliwości i ograniczenia techniki, a z drugiej strony też kontekst społeczny - obowiązki i powinności ludzi do odpowiedniego zachowania się w sytuacjach określonych uwarunkowań wykorzystania nowych technologii.

Wszelkie możliwe scenariusze

Z drugiej strony powinnością twórców rozwiązań technicznych, infrastrukturalnych czy regulacyjnych jest także projektowanie i realizowanie środków zaradczych, by nie utrudniały funkcjonowania ludziom, którzy chcą postępować zgodnie z prawem i normami społecznymi, a jednocześnie móc normalnie, bez zbędnych przeszkód funkcjonować w swoim otoczeniu. M.in. z tego powodu twórcy oprogramowania implementowanego w pojazdach autonomicznych zastanawiają się nad uwzględnieniem jak najlepszych, jak najbardziej przyjaznych człowiekowi a jednocześnie jak najbardziej jednoznacznych i bezpiecznych reakcji na różne, możliwe do zaistnienia w praktyce scenariusze i sytuacje drogowe.

Przykładowo, w obecnych warunkach intuicja człowieka kierującego samochodem podpowiada mu, że jeśli na drogę wpada piłka, to za chwilę wbiegnie za nią dziecko i trzeba zachować szczególną ostrożność. Program sztucznej inteligencji kierujący pojazdem autonomicznym intuicji nie ma, ale trzeba mu zaprogramować działania, które maszyna musi bezwzględnie wykonać w takich i im podobnych sytuacjach. Model musi na przykład uwzględniać, że gdy na drodze pojawi się nagle obiekt nie zaklasyfikowany do przedmiotów związanych z ruchem samochodowym bądź z infrastrukturą drogową, to należy zmniejszyć prędkość, lub zatrzymać się i kontynuować jazdę dopiero po wskazanym czasie, gdy nic więcej się nie wydarzy. Nie mają tu nic do rzeczy etyczne powinności maszyny. Tego rodzaju problemy należy rozwiązywać na poziomie inżynierii oprogramowania a nie w kontekście dywagacji o dylemacie wagonika: czy pojazd ma zabić wybiegające za piłką dziecko, czy też skręcając, aby uniknąć uderzenia, najechać na stojącą po drugiej stronie jezdni jego babcię.

Ten przykład pokazuje, jak ważne jest również dysponowanie odpowiednim zbiorem danych treningowych, które będą uwzględniały jak najwięcej tego typu sytuacji. Co ważne, powinny to być dane pozyskane w jak największym stopniu z lokalnych środowisk, by odzwierciedlały specyficzne, lokalne uwarunkowania różnych zachowań na drodze. Na przykład czas oczekiwania na pojawienie

się dziecka za piłką na osiedlu domów jednorodzinnych może być krótszy niż w pobliżu boiska szkolnego ogrodzonego wysoką siatką i mającego furtkę po drugiej stronie budynku szkolnego.

Na rozpatrywaniu tego rodzaju sytuacji (nawet hipotetycznych) i szukaniu zapobiegających możliwym ryzykom rozwiązań będzie polegało etyczne zachowanie projektantów i wytwórców oraz wdrożeniowców systemów sztucznej inteligencji.

Piąty poziom na horyzoncie

Dostęp do różnych, możliwych scenariuszy zachowań drogowych i rejestrów infrastruktury drogowej w danym miejscu jest niezwykle ważny również ze względu na obecny stan rozwoju transportu autonomicznego. Według organizacji SAE (Society of Automotive Engineers) można wyróżnić pięć poziomów automatyzacji pojazdów. Najwyższy, piąty poziom oznacza całkowitą autonomię działania samochodu. Może on poruszać się w pełni samodzielnie, bez kierowcy, w każdym miejscu i w każdym, dowolnych warunkach drogowych.

Obecnie żaden z wykorzystywanych w przestrzeni publicznej pojazdów nie osiągnął jeszcze piątego poziomu automatyzacji. Są już natomiast auta znajdujące się na czwartym poziomie automatyzacji (na przykład samojazdne taksówki w niektórych miastach USA). Czwarty poziom pozwala na samodzielne poruszanie się samochodu tylko w tzw. domenie projektu operacyjnego (ODD), czyli w ściśle określonym, ograniczonym środowisku drogowym i wskazanych warunkach drogowych (np. przy określonej pogodzie). Przykładowo, przy nagłej zmianie warunków atmosferycznych lub w przypadku opuszczenia wyznaczonej strefy pojazd może się zatrzymać lub poinformować znajdujące się wewnątrz osoby o konieczności przejęcia przez nie kontroli.

W przypadku czwartego poziomu automatyzacji bardzo ważne jest dokładne, wcześniejsze rozpoznanie trasy, po

której będzie się poruszało auto, zmapowanie przestrzeni wokół niego i zidentyfikowanie zarówno rutynowych jak i potencjalnych sytuacji drogowych (np. ryzyk związanych z pierwszeństwem przejazdu na skrzyżowaniach, sposobów działania sygnalizacji świetlnej, rodzajów pojazdów dominujących na drodze). Tego typu rozpoznanie wykonywane jest w różny sposób, albo przez firmy wykorzystujące pojazdy autonomiczne, albo przez funkcjonujących już w niektórych miejscach na rynku dostawców komercyjnych, którzy oferują na przykład zobrazowania lidarowe, albo w ramach projektów publicznych. W Polsce baza scenariuszy testowych dla pojazdów autonomicznych, mająca ułatwić wprowadzanie takich aut na polskie drogi, tworzona jest w ramach projektu DARTS-PL, realizowanego przez konsorcjum złożone z Instytutu Transportu Samochodowego (lider) i Wydziału Elektroniki i Technik Informatycznych Politechniki Warszawskiej.

Nie wiadomo kiedy na drogach pojawią się samochody działające na piątym poziomie automatyzacji. Być może ich wprowadzenie do przestrzeni publicznej będzie wymagało przygotowania odpowiedniego środowiska zapewniającego ich samodzielne funkcjonowanie w każdym, dowolnych warunkach i rozwój systemu transportu z udziałem pojazdów autonomicznych pójdzie w kierunku np. separacji potoków komunikacyjnych.

Specjalizująca się w etyce sztucznej inteligencji prof. Aimee van Wynsberghe twierdzi, że zamiast rozmyślać o programowaniu moralności w maszynach, trzeba skupić się na zapewnieniu ich bezpiecznego działania.

System AI tylko wykonuje powierzone mu zadania – dokonuje obliczeń, których wynik decyduje o podaniu takiej a nie innej odpowiedzi, zaproponowaniu takiego a nie innego rozwiązania. O wynikach postępowania decyduje matematyka a nie wolna wola, świadomość moralna, namysł etyczny, respektowany system wartości, czy inne przynależne ludziom stany i właściwości. Pojazdy autonomiczne nie mogą więc kierować się kryteriami etycznymi. Muszą przede wszystkim działać zgodnie z dobrze zaprojektowanymi przez człowieka regułami technicznymi.