



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

Prof. dr hab. Bogdan Stefanowicz

Wyższa Szkoła Informatyki Stosowanej i Zarządzania
pod auspicjami Polskiej Akademii Nauk

„Sztuczna inteligencja i etyka”

Wstęp

Niezależnie od naszych opinii – pozytywnych czy negatywnych – sztuczna inteligencja umacnia się i rozwija. I nie sposób odwrócić tego procesu. Jest to pewnik, aksjomat. Pozostaje analizować skutki i sprzyjać kierunkom pozytywnym i minimalizować skutki negatywne. To tak jak z każdym wynalazkiem.

A że nie wszystkie skutki obecne i przyszłe wszechobecności sztucznej inteligencji są pozytywne – to rzecz oczywista. Ta technologia to potencjał, który człowiek może wykorzystać rozmaicie: w kierunku pożytku indywidualnego i społecznego, ale także w kierunku niszczącym. Rodzi to konieczność poszukiwania odpowiednich rozwiązań hamujących. Rozwiązań takich można poszukiwać w różnych obszarach. Jednym z nich jest zwrócenie uwagi na głównego aktora – człowieka, który w swoim geniuszu jest często nieobliczalny, jeżeli nie zostanie poddany określonym regułom. Jedne z nich są zewnętrzne, zapisane w kodeksach prawnych; inne – to wewnętrzne wartości, które są silniejsze niż owe zapisy prawne: od prawa można podejmować próby ucieczki, ale od siebie człowiek nie ucieknie.

Reguły wewnętrzne formułuje *etyka*. Jej zasady i reguły są określane jako *strażniczki*, które określają ramy dopuszczalnych zachowań i podejmowanych działań. Zachęca to do poszukiwania takich rozwiązań w obszarze sztucznej inteligencji, które – tak jak w świecie rzeczywistym w odniesieniu do człowieka – będą realizować zadania owych strażniczek w różnych systemach wyposażonych w metody sztucznej inteligencji (SZI).

Aktualna jest teza o konieczności rozwijania koncepcji moralności (etyki) technicznej systemów budowanych na bazie sztucznej inteligencji – i tym bardziej aktualna, im szerzej rozwija się owa sztuczna inteligencja. Artykuł podejmuje pewną dyskusję w tym zakresie.



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

2. Inteligencja – Sztuczna inteligencja

➤ Inteligencja

Inteligencja jest to cecha umysłu regulująca sprawność procesów myślowych, rozwiązywania problemów i innych czynności poznawczych. Cynceron, rzymski pisarz, wybitny mówca, polityk i filozof, przyjął, że są to zdolności umysłowe człowieka. Jan Strelau, polski psycholog, pisał (Strelau, 1987, s. 12): „inteligencja jest właściwością psychiczną, która przejawia się we względnie stałej, charakterystycznej dla jednostki efektywności wykonywania zadań”. Są to zdolności adaptacyjne człowieka do nowych warunków i rozwiązywania nowych zadań. Zachowanie inteligentne różni się od instynktu i prostej reakcji na bodziec ze względu na następujące operacje myślowe:

- przewidywanie – przynajmniej w jakiejś skali czasowej i z pewnym prawdopodobieństwem – przyszłości i skutków podejmowanych działań zanim je podejmiemy,
- formułowanie celów oraz planowanie działań zamierzonych, podejmowanych ze względu na te cele,
- rozumowanie, czyli analiza faktów i wyprowadzanie wniosków na drodze rozważań logicznych (dedukcji) popartych metodami statystycznymi, pozwalającymi uogólniać spostrzeżenia.

Inteligencja to „silnik”, który determinuje intensywność aktywności człowieka. Historia dowodzi, że dzięki tej właściwości człowiek dokonał niezliczonej ilości nader pożytecznych dla ludzkości odkryć i wynalazków. Ale też przyczynił się do wielu tragedii, żeby wymienić tylko bombę atomową.

Zatem silnik taki wymaga steru, który pokieruje człowieka we właściwym kierunku, pożądanym we współczesnym społeczeństwie. Sterem takim jest etyka.

3. Sztuczna inteligencja (SzI)

Termin ten (ang. *Artificial Intelligence*) został wprowadzony przez Johna McCarthy’ego w 1956 r. (początek ery komputerowej!) na pierwszej konferencji poświęconej komputerowej



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

symulacji procesów myślowych. Jest on krytykowany przez wielu specjalistów: co za **sztuczna inteligencja**? Inteligencja jest jedna – naturalna. Ale dotąd stosowany z braku innych, lepszych terminów.

W sprawie tego terminu panuje spór, co on oznacza. Spór ten obejmuje nie tylko środowiska polskie, lecz także środowiska międzynarodowe. W literaturze można spotkać się z trzema różnymi stanowiskami w tej sprawie (jak zresztą w odniesieniu do wielu innych terminów informatycznych):

- 1) Próba jego zdefiniowania.
- 2) Sens tego terminu wyjaśnia się poprzez wskazanie zadań, którymi ma się ona zająć.
- 3) Pozostawia się jego interpretację współrozmówcy: niech się domyśli, „o co nam chodzi”.

Naturalnie, to ostatnie podejście nie świadczy dobrze o osobie, która je stosuje: albo sama nie wie o czym mówi, albo lekceważy rozmówcę.

W artykule roboczo przyjmujemy, że *sztuczna inteligencja* to termin oznaczający pojęcie stosowane w informatyce na określenie sposobów rozwiązywania problemów na wzór naturalnych działań i procesów poznawczych człowieka. Teoretycznie nie wymaga nawet żadnej techniki: wystarczy aparat matematyczny. Współcześnie rozwija się dzięki ogromnym możliwościom techniki komputerowej. Obecnie znane są różne rozwiązania SzI: systemy eksperckie/ekspertowe, algorytmy genetyczne, sieci neuronowe, mobilne technologie agentowe.

Wszystkie one wzorują się w jakimś stopniu na tym, jak człowiek radzi sobie w świecie realnym – wykorzystując potencjał swojej inteligencji. W kierunku pozytywnym, ale też destrukcyjnym. I potencjał ten rośnie wraz z rozwojem owej techniki we wszystkich obszarach ludzkiej aktywności ze wszystkimi skutkami – pozytywnymi i negatywnymi. Naturalnie, należy poszukiwać sposobów wzmocnienia tych pierwszych i minimalizowania drugich.

W minimalizacji (na wyeliminowanie wszystkich w praktyce nie ma co liczyć) skutków negatywnych w życiu człowieka dużą rolę odgrywają wspomniane „strażniczki” – normy etyczne. One to bowiem w jakimś stopniu ograniczają niepożądane negatywne akcje człowieka.

Sugeruje to, że zasadne staje się wbudowanie do wszelkich systemów wyposażonych w sztuczną inteligencję w odpowiednich „sztucznych strażniczek” – modułów modelujących w tych systemach zachowania etyczne.



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

4. Etyka

Nie podejmując szerszych analiz terminu *etyka* przyjmiemy w artykule w sposób nieprofesjonalny, że oznacza on postawę człowieka umożliwiającą mu ustosunkowanie się wobec dobra i zła, jakie mogą wynikać z podejmowanych czynów. Ma pewną moc wiążącą. Etyka to w sensie potocznym — ogół norm moralnych uznawanych przez jakąś zbiorowość społeczną jako punkt odniesienia dla oceny i regulacji postępowania w celu integracji wokół pewnych wartości. Często jest synonimem moralności. Kieruje namysłem nad tym, co i dlaczego w życiu człowieka i społeczeństwa jest dobre oraz co z tego niesprzecznie wynika dla różnych płaszczyzn życia. Wymowne są w tym kontekście słowa Monteskiusza, francuskiego filozofa i pisarza doby Oświecenia, przytoczone przez Stefano Ballerio’go (2019, s. 93): „Gdybym wiedział o czymś, co byłoby dla mnie korzystne, a co byłoby szkodliwe dla mojej rodziny, wyrzuciłbym to z mojego umysłu. Gdybym wiedział o czymś co byłoby korzystne dla mojej rodziny, ale nie dla mojej ojczyzny, starałbym się o tym zapomnieć. Gdybym wiedział o czymś, co byłoby korzystne dla mojej ojczyzny, ale szkodliwe dla Europy, lub też korzystne dla Europy, ale szkodliwe dla rodzaju ludzkiego, uznałbym za zbrodnię”.

Etyka respektuje prawa innych ludzi. Życie według zasad etycznych jest trudne, ale życie bez nich jest nieporównywalnie trudniejsze. *Etyka i inteligencja* to dwie składowe „ustawiające” człowieka w jego relacjach z otaczającą rzeczywistością i innymi członkami tej rzeczywistości.

5. „Sztuczna etyka”

Wspominaliśmy, że – podobnie jak inne rozwiązania, sztuczna inteligencja może być użyta w celach szlachetnych lub złych, szkodliwych. Rodzi to konieczność wdrożenia „strażniczek” strzegących przed jej wykorzystaniem szkodliwym społecznie. Jednym ze sposobów „ucywilizowania” rozwiązań budowanych z wykorzystaniem metod sztucznej inteligencji jest wbudowanie w ich struktury „procedur etycznych”.

Kwestia etyki w sztucznej inteligencji jest już od lat szeroko dyskutowana przez wielu specjalistów. Dowodzą tego wielorakie wypowiedzi w Internecie, w szczególności w kontekście robotyki. Dla przykładu można przytoczyć sformułowane przez Isaaca Asimova, amerykańskiego pisarza, trzy zasady funkcjonowania robotów:



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

1. Robot nie może skrzywdzić człowieka ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy.
2. Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z Pierwszym Prawem.
3. Robot musi chronić samego siebie, o ile tylko nie stoi to w sprzeczności z Pierwszym lub Drugim Prawem.

Przy niewielkich modyfikacjach tych zapisów stosownie do wybranego oprogramowania można uzyskać **gotowe reguły etycznego** zachowania, które mogłyby zostać dopisane do rozwiązań/systemów funkcjonujących na bazie metod sztucznej inteligencji. I choć można mieć różne zastrzeżenia merytoryczne wobec tych zasad (na przykład nie można bezwarunkowo zakładać, że uwzględniają one wszelkie okoliczności), to jednak mogą stanowić zachętę do budowania własnych zasad, dostosowanych do współczesnych wymagań i uwarunkowań.

Z założenia, etyka powstrzymuje człowieka przed podejmowaniem działań prowadzących do szkodliwych skutków. Dotyczy to także jej roli w budowie i funkcjonowaniu wszelkich systemów informacyjnych. Aby etyka była skuteczną „strażniczką”, trzeba zidentyfikować potencjalne zagrożenia. Jest to jedno z kluczowych zadań. Zwraca na to uwagę Julia Bossmann (2016), specjalistka w zakresie sztucznej inteligencji i robotyki. Autorka wymienia m.in. sztuczną głupotę (*artificial stupidity* – podawanie informacji nie na temat), rasizm (*racist robots* – wszelkie programy i systemy budowane przez człowieka zawsze są obciążone jego kryteriami wartościującymi – niekiedy ze złymi intencjami) i złego geniusza (*evil genies* – programy/systemy realizujące żądania, które pociągają nieobliczalne skutki).

6. Wnioski

1. Zarysowują się dwie płaszczyzny związku etyki ze sztuczną inteligencją:
 - Płaszczyzna projektowania rozwiązań z wykorzystaniem tych metod. Wiąże się to z etyką zachowań człowieka: nie narażać siebie i bliźnich na budowanie takich (choćby ze sztuczną inteligencją) systemów, które mogłyby zaszkodzić komukolwiek. W istocie dotyczy to nie tylko rozwiązań ze sztuczną inteligencją, lecz wszelkich systemów rozwijanych na bazie współczesnych technik informatycznych.



**STOWARZYSZENIE INSPEKTORÓW OCHRONY DANYCH
OSOBOWYCH
IV KONFERENCJA SIODO – 25 MAJA 2020**

- Płaszczyzna zasad funkcjonowania budowanych systemów: zasady takie powinny chronić użytkowników przed negatywnymi skutkami działania takich systemów. Ten aspekt można określić jako istota sztucznej etyki. W pierwszym rzędzie trzeba tutaj wymienić wszelkie rozwiązania wykorzystujące technologię mobilnych agentów.
2. Konieczne jest podjęcie badań nad miejscem norm i reguł etycznych w systemach ze sztuczną inteligencją – ze względu na konieczność rozwijania wartości humanistycznych we wszelkiej formie w stechnicyzowanym społeczeństwie informacyjnym. Współczesny imperatyw.
 3. Konieczne jest opracowanie względnie listy zasad/reguł etycznych, które mogłyby być implementowane w takich systemach – zróżnicowanych stosownie do różnych obszarów zastosowań. W szczególności w odniesieniu do wszelkich procesów informacyjnych, jak zbieranie danych, ich przetwarzanie, udostępnianie, ochrona, w procesach edukacyjnych i innych. Interesujące i pożyteczne mogą się okazać propozycje przedstawione przez Bruce'a Weinsteina (2011), autora publikacji z zakresu etyki.
 4. Celowe jest dokonanie wyboru kilku systemów informatycznych i podjęcie eksperymentów w zakresie budowy takich rozwiązań hybrydowych. Na początku nie powinny to być zbyt ambitne plany.

Bibliografia

- Aleksander I. (1988), *Designing Intelligent Systems*. Kogan Page Ltd. 1988.
- Ballerio S. (2019), *Monteskiusz. Ludzie, duch, prawa*. Tłumaczyła Barbara Wierczyńska. Wydawnictwo Hachette Polska.
- Bossmann J. (2016), *Top 9 ethical issues in artificial intelligence*.
<https://www.weforum.org/agenda/2016/10/top-10>
- Bostrom N., Yudkowsky E., *The Ethics of Artificial Intelligence*. MIRI MACHINE INTELLIGENCE RESEARCH INSTITUTE. <https://intelligence.org/files/EthicsofAI.pdf>
- Feigenbaum E., Feldman J. (1972), *Maszyny matematyczne i myślenie*. Tłumaczyła Danuta Gajkovicz. Państwowe Wydawnictwo Naukowe. Warszawa.
- Strelau J. (1987), *O inteligencji człowieka*. Wiedza Powszechna, Warszawa.
- Weinstein B. (2011), *Inteligencja etyczna. Pięć zasad rozwiązywania najtrudniejszych problemów*. Tłumaczenie Magdaleny Szopy. M Wydawnictwo, Kraków.