

Aspekty etyczne sztucznej inteligencji

dr inż. Jacek Martinek

Abstrakt: Dział etyki szczegółowej, zwany *etyką komputerową*, już od lat sześćdziesiątych XX wieku zajmuje się aspektami etycznymi informatyzacji. Według Jamesa Moor'a informatyzacja powoduje powstawanie nowych problemów etycznych, wywołanych przez powszechność komputerów, ich elastyczność logiczną, nieprzejrzystość działania, potencjalną zawodność i możliwość nadużywania. W *Association for Computing Machinery* już w roku 1972 opracowano kodeks etyczno-zawodowy, rozbudowany w roku 1992. Amerykański *The Computer Ethics Institute* (założony w roku 1985) opracował 10 przykazań etyki komputerowej, skierowanych do informatyków i do użytkowników komputerów. Ważnymi źródłami informacji o etyce komputerowej są: książka Wojciecha Bobera, *Powinność w świecie cyfrowym*, wydana w roku 2008 oraz, dostępny w Internecie, zbiór materiałów *Wprowadzenie do etyki informatycznej* opracowany przez Andrzeja Kocikowskiego, Krystynę Górniak-Kocikowską i Terrego Bynuma w roku 2001. W związku z przewidywanym rozwojem robotyki i nanotechnologii od początku XXI wieku rozwija się, tak zwana, *etyka maszyn*, analizująca wymiar etyczny systemów sztucznej inteligencji, aby ich postępowanie wobec ludzi i innych systemów było akceptowalne etycznie. Ostatnio dwa prestiżowe wydawnictwa wydały książki o tytułach *Moral Machines* (Oxford U.P.) i *Robot Ethics* (The MIT Press). Wprowadza się klasyfikację robotów zależnie od ich świadomości moralnej. Do projektowania robotów proponuje się stosować podejście zstępujące, rozpoczynając od przyjęcia zasad moralnych, na których ma opierać się funkcjonowanie agenta, albo podejście wstępujące, w którym rozwijanie moralności następuje przez uczenie na przykładach. Jakościowo nowe problemy etyczne może przynieść rozwój systemów, tak zwanej, *ogólnej sztucznej inteligencji*, w których świadomość etyczna systemu musi być uwzględniana przy jego projektowaniu. Niektóre przyszłe systemy sztucznej inteligencji mogą kandydować do posiadania *statusu moralnego*, analogicznego do statusu moralnego ludzi. W związku z tym na uwagę zasługuje teoria *przyjaznej sztucznej inteligencji* zaproponowana przez Eliezera Yudkowsky'ego.