

Paweł Stacewicz

O INFORMATYCZNEJ INŻYNIERII UMYSŁU

czyli

autorska zachęta do przeczytania książki

UMYSŁ-KOMPUTER-ŚWIAT

O zagadce umysłu z informatycznego punktu widzenia

Pierwsza część książki, do której przeczytania w całości chciałbym zachęcić, nosi tytuł „*Informatyczna inżynieria umysłu*”. W domyśle tytuł ów wieńczy filozoficzny znak zapytania. Cóż bowiem wynika stąd, że inżynier próbuje powierzyć komputerom zadania, które dotychczas podejmował ludzki umysł? Albo inaczej: co współczesne techniki informatyczne, np. te spod znaku sztucznej inteligencji, mogą wnieść do badań i dociekań nad umysłem?

Na przykład:

Czy pozwolą one rozświetlić zagadkę świadomości?

Czy dadzą lepszy wgląd w istotę uczenia się ludzi?

Czy w ogóle umysł da się ująć algorytmicznie i liczbowo?

Próbom rozjaśnienia sensu takich pytań – rozjaśnienia tylko, bo o udzieleniu ostatecznych odpowiedzi nie może być mowy – towarzyszy w książce dwojakiego rodzaju dylemat.

Po pierwsze, jest to dylemat metodologiczny, a mianowicie taki, czy komputery pozwolą tylko *modelować* umysł (zawsze wybiórczo i w sposób niedoskonały), czy też pozwolą go sztucznie *zrealizować*, a więc wytworzyć umysły sztuczne o możliwościach porównywalnych z ludzkimi. Po drugie, jest to dylemat techniczny: czy w obydwu wielkich zadaniach informatycznej inżynierii umysłu, tj. zadaniu modelowania i zadaniu realizacji, lepiej sprawdzą się techniki logiko-podobne (ich dobry przykład stanowią systemy eksperckie), czy też lepsze okażą się metody naturo-podobne (do nich zaliczają się z kolei różnego rodzaju sztuczne sieci neuronowe).

Nakreślone wyżej pytania oraz towarzyszące im dylematy są wyrazem czegoś, co nazywamy w książce informatycznym światopoglądem, krócej zaś: *informatyzmem*. Z grubsza rzecz biorąc, chodzi o filozoficzną nadbudowę nad obecnymi we współczesnej nauce ideami, teoriami i wynikami informatycznymi (a także matematyczno-logicznymi, o ile wiążą się one jakoś z informatyką). Idzie zatem o próbę zbudowania filozofii opartej na informatyce – filozofii krytycznej, bo analizującej wnikliwie możliwości i ograniczenia swych informatycznych fundamentów.

Oto niewielka próbka podejmowanych w książce zagadnień, które wywodząc się z informatyki (lub jej matematycznych podstaw), mają niezwykle ciekawe filozoficzne konsekwencje. Po krótkim anonsie zagadnienia wskazujemy esej, w którym znajduje ono swoje rozwinięcie.

I. Maszyny Turinga

W roku 1936 Alan Turing, angielski matematyk o bardzo inżynierskim podejściu do matematyki, przedstawił ideę abstrakcyjnej maszyny, która miała stanowić kluczowy element negatywnej odpowiedzi na fascynujące pytanie: „Czy wszystkie problemy matematyczne mają rozwiązania algorytmiczne?”. Kilkanaście lat później, gdy powstały już realne maszyny cyfrowe, okazało się, że teoretyczna konstrukcja Turinga opisuje adekwatnie (choć na bardzo podstawowym poziomie) sposób działania każdej maszyny cyfrowej – niezależnie od jej wewnętrznej architektury i stopnia złożoności sterującego nią oprogramowania. Tym samym zaś filozoficzne zagadnienie podobieństwa umysłu i komputera zyskało precyzyjny punkt odniesienia. Można było wyrazić je pytaniem, którego ostatni człon został ściśle zdefiniowany: „Czy umysł jest uniwersalną maszyną Turinga?”.

Więcej o tym w eseju 10, „*Czy umysł jest liczbą?*”

II. Sztuczne sieci neuronowe

W roku 1943 dwaj amerykańscy uczeni, Warren McCulloch (neurofizjolog) oraz Walter Pitts (logik), zaproponowali pierwszy matematyczny model neuronu. Pomysł ten uutorował drogę inżynierom, którzy już niebawem, bo w latach 50-tych XX-ego wieku, zaczęli konstruować układy przetwarzające dane na podobieństwo mózgu. Nazwano je sztucznymi sieciami neuronowymi. Ich współczesne zastosowania są wszechobecne, a obejmują między innymi realizację pewnych prostych czynności poznawczych, jak np. rozpoznawanie kształtów. Oprócz pożytecznych zastosowań idea przetwarzania sieciowego pobudza do niezwykle intensywnej refleksji nad umysłem, którą to kwestię obrazują dobrze dwa pytania: 1) „Jakiego rodzaju sieci modelują adekwatnie wewnątrz-mózgową aktywność neuronów?”, a także 2) „Czy sztuczne sieci neuronowe są czymś jakościowo różnym od maszyny Turinga?”.

Więcej o tym w eseju 4, „*Czy komputer przypomina mózg?*”

III. Automatyzacja uczenia się

W latach 80-tych XX wieku informatycy powołali do życia nową gałąź swojej dyscypliny, nadając jej angielską nazwę „Machine Learning” (po polsku „Automatyzacja uczenia się”). Choć badania wciąż trwają, to ich realny dorobek już obejmuje tysiące algorytmów uczenia się, dzięki którym systemy informatyczne potrafią doskonalić swoje działanie w interakcji ze środowiskiem. Czy algorytmy takie – pomimo licznych odwołań do biologii i psychologii, a także dość dużej skuteczności – odzwierciedlają jednak sposób uczenia się ludzi? Czy w pojęciu „algorytmizacji uczenia się” nie tkwi przypadkiem sprzeczność, pozbawiająca program „Machine Learning” głębszego sensu? Oto pytania, które dotyczą wprost ludzkiego umysłu i naturalnej dlań zdolności do uczenia się.

Więcej o tym w eseju 5, „*Czy roboty powinny się uczyć?*”

IV. Idea nieobliczalności

W latach 30-tych XX-wieku dwaj matematycy, Kurt Gödel i wzmiankowany już Alan Turing, dowiedli (każdy w inny sposób), że istnieją twierdzenia matematyki, których nie da się uzyskać mechanicznie. Owa mechaniczność miała polegać bądź na schematycznym wyprowadzaniu twierdzeń z podanych aksjomatów (tak było u Gödla), bądź na zaprogramowanym uprzednio przekształcaniu symboli przez pewną abstrakcyjną maszynę (tak widział sprawę Turing). Ponieważ zarówno symbole Turinga, jak i formuły Gödla, były kodowane jako liczby, to zaproponowane przez nich procedury trzeba nazwać *obliczeniami* (prowadzącymi do nowych liczb). Stąd też twierdzenia nieosiągalne za pośrednictwem tych procedur zyskały miano *nieobliczalnych*. Zarówno z uwagi na te wyniki, jak i późniejsze prace informatyków, którzy zidentyfikowali wiele prototypowych problemów nieobliczalnych, powstaje pytanie dotyczące tym razem możliwości ludzkiego umysłu: „Jak to jest możliwe, że umysł „widzi” prawdziwość twierdzeń, których maszyna nie potrafi wyprowadzić?”. Albo szerzej: „Jak to się dzieje, że człowiek potrafi rozwiązywać problemy (nieobliczalne!), których rozwiązania są niedostępne dla maszyny (liczącej)?”.

Więcej o tym w eseju 2, „*Czy komputery mogą być nieobliczalne?*”

V. Informatyczne rozwiązanie problemu psycho-fizycznego

Niemal 400 lat temu, w wieku XVII-tym, francuski uczony Rene Descartes stanął wobec problemu, który zdawał się podważać całą jego konstrukcję filozoficzną: w jaki sposób *niematerialna* dusza może oddziaływać na *materialne* ciało, które to oddziaływanie zdaje się być niepodważalnym *faktem*? Próby rozwiązania tego problemu doprowadziły filozofów do różnych fantastycznych koncepcji, jak np. psycho-fizyczny paralelizm czy epifenomenalizm.

Dopiero w wieku XX-tym, gdy powstały automaty sterowane algorytmicznie i wykryzalizowały się na dobre trzy kluczowe pojęcia informatyki (*informacja*, *algorytm* i *automat*), kartezyjski problem znalazł przekonujące wyjaśnienie. Jest to wyjaśnienie przez analogię, któremu siłą daje pojęcie informacji. Wiemy bowiem, że zakodowana w algorytmach informacja jest, z jednej strony, czymś materialnym (ma swój nośnik fizyczny), a z drugiej strony czymś abstrakcyjnym (definiującym pewien schemat działania). Wiemy nadto, że owe abstrakcyjne schematy mogą być przyczyną fizycznej interakcji maszyny ze środowiskiem. Wiedza ta nasuwa z kolei tezę, że umysł niczym system przetwarzający informacje odbiera z zewnątrz pewne dane, przekształca je w oparciu o wewnętrzne algorytmy i na podstawie uzyskanych wyników kieruje zachowaniem cielesnego automatu. W ten sposób problem psycho-fizyczny znika, a unicestwia go „materialno-niematerialne” pojęcie informacji.

Więcej o tym w eseju 9, „*Czy komputery przetwarzają informację?*”

Gorąco zachęcam do lektury

Paweł Stacewicz