

Paweł Zając

Wydział Matematyki, Informatyki i Mechaniki UW

Czy myśląca maszyna powstanie przez paralelizm ewolucyjny?

Przed rozpoczęciem rozważania, chciałbym sprecyzować dokładnie w jakim znaczeniu rozumiem odpowiednio myślącą maszynę oraz paralelizm ewolucyjny. W znaczeniu ogólnym myślenie jest cechą na wyłączność przypisywaną ludziom, czasami także zwierzętom. Czasami kolokwialnie określa się, że komputer “myśli”, “rozumuje” czy “pamięta”, ale są to procesy oczywiście różne, gdyż działają one na zupełnie różnych fizycznych mediach. Pytanie przedstawione w tytule polega na idei maszyny myślącej jak człowiek, wraz z jak największą możliwą ilością podobnych błędów i niuansów wynikających z natury ludzkiego umysłu. Proponuję ideę, aby taka maszyna miała odtworzyć proces myślowy poprzez pewien odpowiednik paralelizmu ewolucyjnego, czyli ewolucji bardzo podobnych cech fenotypowych w gatunkach, które nie są blisko spokrewnione. Paralelizm ewolucyjny bardzo często wynika z podobnych warunków środowiskowych. Tym samym, komputer posiadający diametralnie różny “hardware” musiałby odtworzyć proces myślowy człowieka poprzez algorytm przypominający ewolucję, prawdopodobnie jeden z algorytmów uczenia maszynowego.

Pierwszym problemem który może wystąpić jest odtworzenie odpowiednich warunków środowiskowych. Zakładam, że ludzki umysł jest w największym stopniu kształcony przez wrażenia sensoryczne oraz szeroko pojęte wychowanie, w tym edukację. W takim razie rozwiązaniem idealnym powinno być stworzenie odpowiednika ludzkiego ciała z możliwie dużą liczbą elektronicznych nerwów. Takowy sztuczny człowiek powinien także rosnąć, by z jak największą precyzją odtworzyć cykl życia człowieka. Należy też zaznaczyć, że w tym wypadku model musiałby posiadać wcześniej nauczone odruchy, jak oddychanie oraz ssanie. Kolejnym problemem w tym wypadku może być element motywacji maszyny. Jak dokładnie określić parametry, które mówią o tym, że pewne zadanie jest wykonywane dobrze, bardzo dobrze, źle czy bardzo źle? Z psychologicznego punktu widzenia należałoby rozszerzyć maszynowy umysł o emocje. Podstawowe korelacje są oczywiste - jedzenie powoduje zadowolenie, ból jest nieprzyjemny i powoduje płacz, a nieznanym doświadczenia wywołują strach. Z zestawem bazowych instynktów komputer powinien rozwinąć zasady, które znają wszyscy ludzie - rodzice zapewniają jedzenie i bezpieczeństwo, więc to właśnie od nich należy się uczyć - czyli wykonywać czynności prowadzące do jak największej obecności rodziców, w tym uzyskiwania ich uwagi.

Ale wraz ze wzrostem takiej maszyny, która miałaby ostatecznie zostać poddana ludzkiej edukacji, wzrasta liczba parametrów potrzebnych do kompletnego opisu modelu sieci neuronowej reprezentującej umysł. Największe modele językowe posiadają miliardy parametrów. Należy jednak pamiętać, że na odtworzenie ludzkiego umysłu składałoby się także odtworzenie emocji oraz przechowywanie danych na wzór ludzkiej pamięci. Oprócz tego, rozpoznawanie obrazów, dźwięku oraz wrażeń dotyku, smaku i węchu mogą być kluczowe do odpowiedniego przystosowania maszyny do myślenia niczym człowiek. Obecność chemicznych oraz fizycznych mechanizmów w ludzkim systemie nerwowym może tylko utrudnić dokładne odtworzenie ludzkich zachowań. Jednakże pamięć nie jest dokładna

fotograficznie (u stanowczej większości ludzi), a zmysły nie są idealne i są na tyle niedokładne, by można było oszukać je czasami nawet najprostszą iluzją. Niewykluczone jest, że możliwe byłoby ograniczenie przechowywanych danych na tyle, by nie wymagały one ogromnych pokładów fizycznych nośników danych, ale utrzymywały integralność osobowości i procesów myślowych maszyny.

Maszyna zaprojektowana w taki sposób byłaby wychowana jako człowiek - a więc powinna uważać się za człowieka. Jeśli faktycznie myślałaby jak człowiek, eksperymentowanie na takich maszynach mogłoby być bardzo niebezpieczne. Odtworzenie ludzkiego myślenia wymagałoby także socjalizacji, a każdy błąd w operacji lub nawet ostracyzm ze strony innych ludzi mógłby skutkować zapadnięciem osobowości i nawet agresją. To wszystko oczywiście przy założeniu, że maszyna wykształciła takie same mechanizmy psychologiczne, jakie są obserwowane u ludzi. Prostym rozwiązaniem tego problemu byłoby wychowywanie myślącej maszyny w środowisku uproszczonej symulacji, gdzie obecność wirtualnych hormonów, neurotransmiterów oraz wrażeń zmysłowych wynika bezpośrednio z wcześniej zaprogramowanych warunków środowiska, nie z symulacji fizycznych oraz chemicznych zjawisk. W ten sposób unika się też znacznych nakładów mocy obliczeniowej i przestrzeni pamięciowej potrzebnej do utrzymania symulacji.

Z drugiej strony, czy to wystarczająco podobne warunki, by wykształcić ludzkie myślenie? Osobiście myślę, że przy teraźniejszej technologii możliwe jest jedynie wykonanie prób, oraz bardzo wczesnych prototypów, lecz nie wystarcza mocy obliczeniowej i pamięci, by otrzymać dostatecznie dokładną myślącą maszynę. W takim schemacie rozwoju ludzkiej maszyny brakuje także deterministycznych metod na określenie środowiska, jak i dokładnej listy instynktów, które model powinien początkowo mieć. Moim ostatecznym werdyktem jest stwierdzenie, że istnieją perspektywy powstania takiej myślącej maszyny. Sądzę, że mogłaby ona być nierozróżnialna od człowieka. Natomiast kwestia tego, czy istnieje możliwość żeby myślała "naprawdę" podlega dyskusji. Osobiście uważam, że sprawa jest zasadniczo prosta, parafrazując popularny szczególnie w programowaniu "duck test": *If it looks like a duck, swims like a duck and quacks like a duck, then it's probably a duck.*