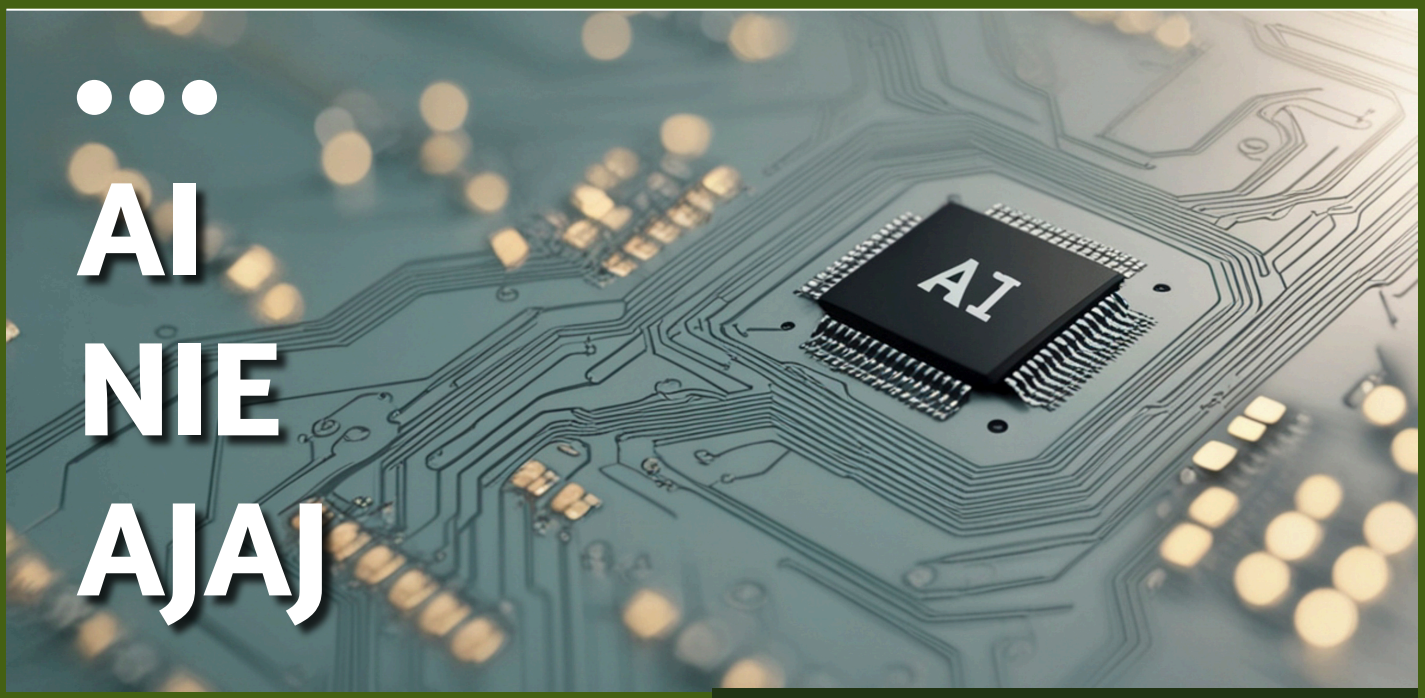

... AI NIE AJAJ

Gdy dziś słyszymy „sztuczna inteligencja", przed oczami stają nam chatboty, generatory obrazów, algorytmy rekomendacji i modele, które potrafią pisać teksty, streszczać książki, tworzyć kod, a czasem nawet zaskakująco dobrze udawać rozmowę z człowiekiem. Łatwo wtedy uwierzyć, że to wynalazek „ostatnich kilku lat". Tymczasem AI jest raczej długą historią marzenia o maszynie, która nie tylko liczy, ale też „rozumuje" – i historią równie długich rozczarowań, okresów zachwytu, zimnych pryszniców oraz powrotów na fali nowych danych i nowej mocy obliczeniowej.

Warto zacząć od rzeczy pozornie banalnej: zanim powstały komputery, najpierw musiał powstać pomysł, że myślenie da się w ogóle przełożyć na procedury. W XVII wieku pojawiają się mechaniczne kalkulatory – urządzenia, które w istocie robią jedną rzecz, ale robią ją konsekwentnie: formalizują rachunek. Pascaline, skonstruowana przez Blaise'a Pascala w latach 1642–1644, potrafiła dodawać i odejmować, a w praktyce była odpowiedzią na bardzo „ziemskie" potrzeby: rozliczenia, podatki, biurokrację. Kilkadziesiąt lat później Leibniz projektuje Step Reckoner (kalkulator wykorzystujący bęben schodkowy), który rozszerza repertuar o mnożenie i dzielenie. To jeszcze nie jest inteligencja, ale jest coś ważnego: przekonanie, że część czynności umysłowych można zamknąć w mechanizmie.

W XIX wieku Charles Babbage idzie o krok dalej – od „maszyny do rachunków" do „maszyny ogólnego przeznaczenia". Jego Analytical Engine, opisany po raz pierwszy w 1837 roku, był projektem mechanicznego komputera, który miał pamięć („Store"), jednostkę wykonawczą („Mill") i mógł być programowany kartami perforowanymi.

To moment, w którym rodzi się intuicja fundamentalna dla późniejszej informatyki i AI: ta sama maszyna może wykonywać różne zadania, jeśli tylko damy jej odpowiedni program.

Skok w XX wiek przynosi formalną rewolucję. Alan Turing w 1936 roku publikuje pracę „On Computable Numbers…", w której pokazuje, jak opisać obliczenia w sposób abstrakcyjny, oraz wprowadza ideę maszyny uniwersalnej – urządzenia, które potrafi naśladować działanie dowolnej innej maszyny tego typu, jeśli dostanie jej opis. To w pewnym sensie „matematyczny przepis" na komputer. A kiedy komputer staje się ideą, a nie tylko urządzeniem, pojawia się naturalne pytanie: czy da się nim imitować także inteligencję?

W 1950 roku Turing publikuje słynny esej „Computing Machinery and Intelligence", zaczynając od prowokacji „Czy maszyny mogą myśleć?" i proponując zamiast sporów o definicje – test rozmowy, dziś znany jako Test Turinga (a u niego jako „gra w imitację"). W tym samym okresie rodzi się środowisko badaczy, którzy zaczynają traktować „inteligencję" jako problem inżynierski. Symbolem narodzin AI jako dziedziny jest projekt warsztatów w Dartmouth: w propozycji z 1955 roku John McCarthy i współautorzy używają terminu „artificial intelligence" i zakładają, że wiele aspektów uczenia się i inteligencji da się opisać tak precyzyjnie, by maszyna mogła je symulować. W tym miejscu historia zaczyna się rozgałęziać na dwa nurty, które będą się ścierały przez dekady. Pierwszy to AI symboliczna: reguły, logika, reprezentacje wiedzy, „jeśli–to", systemy wnioskujące. .

Drugi to nurt inspirowany mózgiem: sieci neuronowe i uczenie się z danych. Co ciekawe, oba nurty bardzo wcześnie dostają swoje „mity założycielskie" – i oba równie wcześnie zderzają się ze ścianą.

Z jednej strony mamy wiarę w reguły i dialog. W 1966 roku Joseph Weizenbaum publikuje opis ELIZY – programu, który symuluje rozmowę, stosując dopasowywanie wzorców i proste transformacje zdań. Eliza nie rozumie, nie ma modelu świata, ale potrafi „brzmieć" jak terapeuta w stylu Rogera: dopytuje, odbija pytania, parafrazuje. Efekt psychologiczny był (i jest) zaskakująco silny: ludzie mają tendencję dopisywać maszynie intencje i głębię, jeśli tylko odpowiedzi pasują do oczekiwań rozmowy. To ważna lekcja również dziś: im bardziej interfejs przypomina dialog, tym łatwiej uwierzyć w „zrozumienie", nawet jeśli pod spodem jest statystyka i sprytna sztuczka.

Z drugiej strony, równolegle rozwijają się idee uczenia. Frank Rosenblatt w 1958 roku opisuje perceptron – model, który miał być uproszczoną matematyczną analogią uczenia się w mózgu. Entuzjazm był ogromny, ale pod koniec lat 60. przychodzi mocna krytyka: Minsky i Papert pokazują ograniczenia prostych perceptronów, co (razem z innymi czynnikami) chłodzi zainteresowanie sieciami neuronowymi na lata. Z kolei w obszarze przetwarzania języka naturalnego ważnym zimnym prysznicem staje się raport ALPAC z 1966 roku, który krytycznie ocenia ówczesne próby tłumaczenia maszynowego i skutkuje ograniczeniem finansowania.

To jest pierwszy wielki motyw historii AI: fala obietnic, po której przychodzi fala rozczarowań. W latach 70. dochodzi kolejny symboliczny cios w Wielkiej Brytanii: tzw. raport Lighthilla (1973) ocenia stan badań AI sceptycznie i staje się jednym z elementów prowadzących do ograniczenia wsparcia dla tej dziedziny w części ośrodków. Mówi się wtedy o „AI winter" – zimie sztucznej inteligencji, czyli okresie, w którym hype przestaje się spinać z rzeczywistością.

A jednak – i to jest drugi wielki motyw – AI nie tyle „umiera", co zmienia formę. Kiedy nie wychodzi „ogólna inteligencja", świetnie zaczynają działać rozwiązania wąskie, specjalistyczne. Tak rodzi się era systemów eksperckich. DENDRAL (już w latach 60., rozwijany dalej) bywa opisywany jako jeden z pierwszych ekspertów komputerowych: system wykorzystujący wiedzę dziedzinową do formułowania hipotez w chemii. W latach 70. pojawia się MYCIN – program doradzający w leczeniu infekcji, oparty o reguły i mechanizm wnioskowania, kojarzony z pionierską falą ekspertów medycznych.

Systemy eksperckie były czymś więcej niż demonstracją: działały w praktyce, w ograniczonych domenach, i sprawiały wrażenie „myślących". Ale miały też wrodzoną kruchość. Po pierwsze: wymagały żmudnego „wbijania" wiedzy przez ekspertów, co nazywa się czasem bottleneckiem akwizycji wiedzy. Po drugie: świetnie radziły sobie w idealnych, przewidzianych scenariuszach, a potrafiły się wykoleić przy małej zmianie kontekstu. Świat realny jest brudny, pełen wyjątków, a „reguł" trzeba nieskończenie wiele. W latach 80. i na początku 90. kolejna fala przesadnych oczekiwań pęka, a AI znów traci część blasku.

Przełom przychodzi z innej strony: zamiast ręcznie programować inteligencję, zaczynamy ją „wyciągać" z danych. To jest moment, w którym uczenie maszynowe (machine learning) staje się centrum ciężkości AI. Istotnym elementem tej zmiany jest powrót sieci neuronowych – tym razem dzięki matematyce i mocy obliczeniowej. W 1986 roku Rumelhart, Hinton i Williams publikują opis skutecznej procedury uczenia sieci wielowarstwowych, znanej jako wsteczna propagacja błędu (backpropagation). To jeszcze nie jest „boom", ale to fundament, na którym później stanie głębokie uczenie.

Równolegle rosną metody statystyczne: modele probabilistyczne, klasyfikatory, algorytmy uczące się z przykładów. W wielu zastosowaniach (rozpoznawanie mowy, analiza tekstu, wykrywanie spamu) nie trzeba „rozumienia" w ludzkim sensie – wystarczy dobrze uchwycić regularności w danych. I właśnie wtedy pojawiają się symboliczne momenty, które działają na wyobraźnię mediów.

W 1997 roku IBM-owski Deep Blue pokonuje Garry'ego Kasparowa w meczu rozegranym w warunkach turniejowych. To nie była „ludzka inteligencja w maszynie", raczej brutalna siła obliczeń połączona z doskonałą inżynierią heurystyk i oceny pozycji, ale przekaz był prosty: komputer wygrał z człowiekiem w grze-symbolu intelektu. W 2016 roku podobny efekt medialny daje AlphaGo, który w meczu w Seulu pokonuje Lee Sedola 4–1, a praca w „Nature" opisuje połączenie sieci neuronowych (policy i value networks) z przeszukiwaniem drzew i uczeniem ze wzmocnieniem.

Ale największa zmiana nie polegała na wygrywaniu gier. Polegała na tym, że AI zaczęła „widzieć" i „słyszeć" lepiej, bo nauczyła się reprezentacji. W 2012 roku model znany jako AlexNet wygrywa konkurs ImageNet, pokazując, że głębokie sieci konwolucyjne potrafią skokowo poprawić rozpoznawanie obrazów, jeśli dostaną duże dane i odpowiedni trening na GPU. Od tego momentu deep learning staje się lokomotywą wielu dziedzin: od wizji komputerowej, przez mowę, po analizę tekstu.

I wtedy dochodzimy do etapu, który w Twoim poście domyka klamrą: współczesnych chatbotów i modeli językowych. Tu kluczowa okazuje się architektura transformera, opisana w 2017 roku w pracy „Attention Is All You Need". Transformer nie jest „mózgiem" i nie jest „zbiorem reguł". To bardzo skuteczny sposób uczenia się zależności w sekwencjach (np. w zdaniach), oparty o mechanizm uwagi (attention), który pozwala modelowi ważyć, które fragmenty kontekstu są istotne. Na tej idei wyrastają modele językowe, które uczą się przewidywać kolejne słowo (a dokładniej: kolejny token) na podstawie ogromnych zbiorów tekstów. Z pozoru to skromne zadanie, ale konsekwencje są nieintuicyjne: jeśli dobrze potrafisz przewidywać, jak ludzie kończą zdania, to zaczynasz też nieźle „udawać" rozumienie.

W tym punkcie wracamy do Testu Turinga – tylko w wersji masowej i codziennej. OpenAI opublikowało wpis „Introducing ChatGPT" 30 listopada 2022 roku, przedstawiając model dialogowy jako „research preview" i zapraszając do testów. I tu następuje społeczny przełom: AI przestaje być narzędziem w tle, a staje się interfejsem. Zamiast klikać w opcje, rozmawiamy. Zamiast uczyć się składni zapytań, piszemy po ludzku. Zamiast „wyszukiwania" – prosimy o odpowiedź.

Warto jednak dodać ważną, często pomijaną rzecz: modele językowe nie „wiedzą" w takim sensie, jak my. One wytwarzają najbardziej prawdopodobny ciąg znaków w danym kontekście. Dlatego potrafią brzmieć pewnie, nawet gdy się mylą. OpenAI wprost ostrzega, że ChatGPT „czasem pisze odpowiedzi brzmiące wiarygodnie, ale niepoprawne albo bezsensowne". To zdanie jest być może jednym z najbardziej praktycznych podsumowań współczesnej AI: niesamowicie użyteczna, ale wymagająca krytycznego myślenia i weryfikacji.

Druga strona tej historii to koszty – nie tylko finansowe, ale też energetyczne i środowiskowe. W miarę jak modele rosną, rośnie też apetyt na obliczenia. Międzynarodowa Agencja Energetyczna (IEA) szacuje, że centra danych zużyły w 2024 roku około 415 TWh energii elektrycznej (ok. 1,5% globalnej konsumpcji) i w scenariuszu bazowym ich zużycie może wzrosnąć do ok. 945 TWh do 2030 roku, częściowo napędzane przez serwery akcelerowane wykorzystywane w AI. MIT w analizie wpływu generatywnej AI podkreśla, że rosnące zużycie energii centrów danych staje się istotnym problemem infrastrukturalnym, nawet jeśli nie cała ta energia idzie stricte na modele generatywne. A to prowadzi do bardzo współczesnego napięcia: chcemy inteligentnych narzędzi „zawsze pod ręką", ale „zawsze pod ręką" oznacza hale serwerów, chłodzenie, prąd, wodę, sieć i ogromny łańcuch dostaw. Jest jeszcze trzeci koszt: informacyjny. Skoro AI potrafi generować tekst, obraz i dźwięk, to świat może zostać zalany treściami syntetycznymi. To nie musi być złe – automatyzacja może uwolnić czas, wesprzeć osoby z niepełnosprawnościami, przyspieszyć naukę i pracę.

Ale może też utrudnić odróżnianie wiarygodnych źródeł od „ładnie brzmiącej fikcji", zwłaszcza gdy treści są produkowane szybko i tanio. W praktyce rośnie znaczenie higieny informacyjnej: sprawdzania źródeł, weryfikowania liczb, odróżniania opinii od faktów, rozumienia, kiedy model może „zgadywać".

Jeśli więc spojrzeć na tę historię z dystansu, zobaczymy coś więcej niż linię od kalkulatora do chatbota. Zobaczymy ruch wahadła między dwiema pokusami: żeby opisać inteligencję jako zestaw reguł oraz żeby „wyhodować" ją z danych. Zobaczymy powtarzający się rytm: nadzieja, hype, rozczarowanie, a potem powrót w nowej formie, gdy technologia dogoni ambicje. Zobaczymy też, że AI jest odbiciem naszych ograniczeń: kiedy brakuje danych – model nie ma z czego się uczyć; kiedy brakuje energii i infrastruktury – skala nie rośnie; kiedy brakuje krytycznego myślenia – nawet najlepsze narzędzie staje się źródłem błędów.

I wreszcie zobaczymy coś bardzo ludzkiego: że tak jak w czasach Elizy łatwo było poczuć, że „ktoś nas rozumie", tak dziś łatwo pomylić płynną odpowiedź z prawdą. Historia AI nie jest więc tylko historią maszyn. To także historia nas – naszej fascynacji inteligencją, naszej skłonności do antropomorfizacji i naszej odpowiedzialności za to, jak tej technologii użyjemy.