

Czy algorytm może przemówić w imieniu sprawiedliwości?

*Językowe wyzwania i perspektywy sztucznej inteligencji w sądownictwie**

Na początek przybliżę pewien kontekst akademicki, który ma niebagatelny wpływ na perspektywę z której podchodzę do tematu. Obecnie jestem u kresu swojej niesamowitej podróży w Szkole Doktorskiej w dyscyplinie nauk prawnych. Jednak w swojej karierze akademickiej odbyłem wspaniałą, pięcioletnią przygodę studiując prawo i równie ważną dla mnie drugą pięcioletnią przygodę studiując matematykę. Moim głównym obszarem badań jest kryminalistyka, ale pobocznie romansuję też z logiką formalną i matematycznymi podstawami sztucznej inteligencji. W zawodowej, pozauniwersyteckiej aktywności projektuję, buduję i stosuję systemy AI.

Pojęcie języka jeszcze do niedawna było dla mnie jasne. Funkcjonując w tradycji koncepcji, które przekazał mi mój mistrz (matematyk prof. Świrydowicz), które to przekazał mu jego mistrz (prawnik prof. Ziemiński), które to opracował jego mistrz (filozof prof. Ajdukiewicz), językiem było dla mnie wszystko co powstawało i funkcjonowało na podstawie: 1) reguł wyznaczających słownik, 2) reguł gramatycznych, 3) reguł aksjomatycznych, 4) reguł inferencyjnych, 5) reguł ustalających uniwersum i 6) reguł denotowania**.

Taka formalna definicja jest wygodna dla prawników, matematyków czy logików, bo czyni z języka narzędzie do osiągania konkretnych celów i rezonuje z naszym zamiłowaniem do zamykania rzeczywistości w sztywne ramy. Można powiedzieć, że takie spojrzenie dawało (przynajmniej mnie) swoiste poczucie bezpieczeństwa wyrastające z przekonania o jasności zagadnienia.

Jednakże, pojawienie się ChatGPT, zwłaszcza modelu GPT-4o który odpowiada jak człowiek, zmusiło mnie do zrewidowania swojego spojrzenia na język.

Skrót GPT pochodzi od „Generative Pre-trained Transformer”. Modele z tej rodziny przekształcają (tudzież transformują) jeden tekst w inny. To wszystko, nic więcej, nic mniej. Modele GPT funkcjonują na podstawie skomplikowanych sieci neuronowych, które wbrew obiegowej opinii mają niewiele wspólnego ze strukturami w naszym mózgu. W porządku, koncepcyjnie są w pewnym stopniu zbliżone, bowiem głównym mechanizmem obu jest skomplikowana sieć połączeń. Tu jednak podobieństwa się kończą. Zaczniemy od tego, że struktura neuronów w mózgu jest fizyczną konstrukcją przetwarzającą sygnały elektorchemiczne. Dosłownie, potencjalnie moglibyśmy rozciąć człowiekowi czaszkę i policzyć neurony. (Nie polecam, głównie dlatego że mikroskopy są drogie). Osobiście widziałem neuron na zajęciach z biologii w gimnazjum (podejrzewam, że nie był ludzki, ale jednak neuron). A jak jest z komputerowymi sieciami neuronowymi w modelach językowych? Otóż, nikt ich nie widział. Więcej, one nie istnieją. Są tylko konstrukcją wirtualną, swoistą instrukcją mówiącą jakie i w jakiej kolejności obliczenia ma wykonać procesor komputera.

Komputerowe sieci neuronowe nie przetwarzają bodźców zewnętrznych i wewnętrznych jak nasze mózgi. Generują tylko najbardziej prawdopodobny ciąg znaków na podstawie danych, które były im dostarczane w momencie konstrukcji modelu (tzw. dane uczące) i w momencie korzystania z modelu (tzw. input). Wobec tego, proces rządzący biologiczną siecią neuronową i proces rządzący komputerową siecią neuronową to dwa istotnie i jakościowo różne procesy.

Gdzie najłatwiej dostrzec tę różnicę? Kartezjusz zaproponował koncepcję skrajnego dualizmu psychofizycznego. Stanowisko Kartezjusza prawdopodobnie było zbyt skrajne, daje jednak ciekawy ogląd na sytuację. Kartezjusz wyróżnił dwie niezależne od siebie substancje: materialną i duchową. Substancja duchowa charakteryzuje się tym, że wątpi, rozumie, pojmuję, twierdzi, przeczy, chce, nie chce, wyobraża sobie i czuje. Tęże substancję (w postaciach mniej lub bardziej skrajnych) psychologowie zwykli nazywać duszą lub umysłem. Żaden z wymienionych atrybutów umysłu nie jest domeną modelu językowego.

Dochodzimy teraz do pytania, czy to co generuje model językowy, faktycznie jest językiem? Według ajdukiewiczowskiej formalnej definicji języka jak najbardziej tak. Z tym, że przyjmowanie tego stanowiska stanowi tak naprawdę zubożenie koncepcji języka lub nawet człowieka, bo – w terminach arystotelesowskich – stanowi obdarcie materii z jej formy. Istotą języka powinno być przekazywanie własności i stanów umysłu. W przeciwnym razie otrzymujemy twór, który wygląda jak język, który działa jak język, ale jest czymś na zasadzie quasi-języka. Sytuacja nie ulegnie poprawie, bowiem AI oparte na sieciach neuronowych nigdy nie posiadzie umysłu.

Brak umysłu i generowanie najbardziej prawdopodobnego ciągu znaków niesie ze sobą poważne zagrożenie dla sądownictwa. Otóż jednym z najchętniej rozważanych przez sędziów zastosowań AI jest uproszczenie przygotowywania uzasadnień wyroków sądowych. Zadania żmudnego i nielubianego przez sędziów.

Nie wyobrażam sobie tego. Model językowy jest w stanie uzasadnić wszystko (jakość takiego uzasadnienia to odrębna sprawa). Bez problemu odpowiednimi promptami możemy uzyskać uzasadnienia dla sprzecznych względem siebie twierdzeń. Wobec tego mogę podjąć arbitralną decyzję i za pomocą modelu językowego bez wysiłku otrzymać dla niej uzasadnienie. W mojej ocenie cienka linia dzieli tę wizję od totalitaryzmu.

Uzasadnienie wyroku sądowego musi być językowym odzwierciedleniem procesów mentalnych sędziego. Decyzję procesową musi poprzedzać analiza norm prawnych, analiza stanu faktycznego oraz subsumpcja. Przebieg tegoż procesu musi znaleźć odzwierciedlenie w uzasadnieniu, które ma zapobiegać arbitralności sądów. Co ochroni nas przed despotcją jeśli model językowy zacznie stawiać fasadę generując uzasadnienia dowolnych twierdzeń?

W takim razie, może w całości oddajmy kompetencje sądownicze modelom językowym, eliminując z procesu skłonnych do nadużyć ludzi? Hell no! Zdaję sobie sprawę, że jakość modeli językowych rośnie, pamiętajmy że model językowy na zawsze pozostanie tylko generatorem najbardziej prawdopodobnego tekstu.

A czy odpowiedź najbardziej prawdopodobna jest jednocześnie odpowiedzią najbardziej słuszną?

Więcej! Funkcjonalny model językowy musiał zostać wytrenowany na ogromnej liczbie danych tekstowych. Danych które pochodzą z różnych źródeł, głównie z Internetu, nie oszukujmy się. Jaki procent z ogólnodostępnego tekstu stanowią teksty prawnicze? Jaki procent z ogólnodostępnego tekstu stanowią teksty akademickie? Kropla w morzu. Przekazanie jakiegokolwiek władzy modelom językowym niewiele różniłoby się od stawiania celebryty z wiedzą potoczną w roli eksperta. Nie idźmy tą drogą.

Rosnąca popularność AI tworzy jeszcze dwa kolejne zagrożenia. Po pierwsze, poprawianie jakości modelu wymaga uczenia go na coraz większych zbiorach danych. Nieuniknionym jest, że niedługo modele językowe w zbiorach danych uczących będą zawierały teksty wygenerowane przez AI. A niestety, uczenie modelu na danych wygenerowanych przez model skutkuje nieuchronnym spadkiem jego jakości. W konsekwencji, fenomen AI jest raczej węzłem zjadającym własny ogon.

Po drugie, coraz większe modele językowe będą wymagały coraz większych mocy obliczeniowych. Uruchamianie AI na urządzeniach lokalnych już teraz stanowi wyzwanie i korzysta się raczej z infrastruktur chmurowych. Ta sytuacja nieprędko ulegnie poprawie, jeśli w ogóle. Oznacza to, że korzystanie z AI wymaga wysłania danych do zewnętrznego dostawcy i utraty wszelkiej kontroli nad tym co z tymi danymi się dzieje. Stwarza to ogromne problemy na gruncie ochrony danych osobowych, ale to temat na zupełnie inną okazję.

Nie chciałbym zostać odebrany jako przeciwnik AI, bo absolutnie nie twierdzę, że prawnicy i modele językowe na zawsze powinni pozostać zwaśnionymi obozami. Język jest dla prawników narzędziem pracy i obiektem badań. W efekcie nie wszystkie zastosowania języka służą prawnikom do intersubiektywnej komunikacji ich wewnętrznych procesów myślowych. Czysto instrumentalne wykorzystywanie języka też się zdarza i to wcale nierzadko. Zwróćmy uwagę choćby na notariuszy. Współczesny rejent jest strażnikiem interesu stron postępowania notarialnego i ma zapobiegać podejmowaniu pochopnych decyzji w sprawach uznanych przez ustawodawcę za naprawdę ważne. Czy wygenerowanie projektu aktu notarialnego przez model językowy w jakiś sposób ujmuje tym zadaniom? Notariusz zamiast pisać całość dokumentu mógłby go tylko skontrolować i zatwierdzić. Zastosowanie tu modelu językowego, który potencjalnie jest szybszy i bardziej niezawodny od ludzi mógłoby przyczynić się do poprawy usług notarialnych. Zwłaszcza że obecna praktyka jest dość podobna, gdzie wczesny projekt aktu notarialnego jest przygotowany przez aplikantów lub sekretariat.

Odchodząc od notariuszy spójrzmy szerzej. Prawnicy wszystkich specjalności często mierzą się z wyzwaniem zapoznania się z ogromną ilością tekstu, jak np. dokumentów dotyczących stanu faktycznego, umów, czy aktów prawnych. Często duża część tych tekstów nie jest istotna dla sprawy, a przeszukiwanie ich stanowi wyzwanie. System AI wykazujący się semantycznym rozumieniem tekstu i generującym dowolne formy podsumowania, dowolnie wskazanych fragmentów ma potencjał istotnego usprawnienia pracy prawnika.

Czy w takim razie mamy jakąkolwiek perspektywę na zastąpienie prawników przez AI w pewnych zadaniach? Oczywiście że tak. Jedną z dotychczas ekskluzywnych dla prawników umiejętności było sprawne poruszanie się po systemie prawnym i rozumienie przepisów. Gdyby taką umiejętność posiadali normalni ludzie, znaczna liczba prostych problemów mogłaby zostać z sukcesem rozwiązana bez udziału prawnika. Obecnie istnieje wyspecjalizowany, polski system AI – Legitio.pl – który z powodzeniem może pomóc w prostej sprawie. Legitio potrafi analizować proste problemy i dostarczać porady na tyle dobrej jakości by były wystarczające w większości spraw życia codziennego. Jednakże, należy pamiętać że tylko człowiek potrafi ocenić czy klient pominął istotne szczegóły sprawy i ewentualnie o nie dopytać. Ta zdolność (a może obowiązek) jest powodem dla którego tylko prawnik-człowiek może w pełni odpowiadać za udzielane porady.

Nie wiem czy AI kiedykolwiek zastąpi prawników, ale w mojej ocenie udzielanie porad prawnych jest najbardziej zagrożonym obszarem ich praktyki.

*referat przygotowany na III ogólnopolską konferencję naukową z cyklu CZŁOWIEK – JĘZYK – PRAWO, pt. „Prawnie – (nie)poprawnie”, Poznań 18-19. listopada 2024r

**Reguły wyznaczające słownik określają jakie słowa należą do danego języka. Reguły gramatyczne określają jak ze słów konstruować wyrażenia bardziej złożone. Reguły aksjomatyczne wyróżniają pewne zdania jako prawdziwe niezależnie od innych zdań. Reguły inferencyjne wskazują jakie zdania języka są prawdziwe na podstawie prawdziwości innych zdań. Reguły denotowania określają do czego odnosi się dane wyrażenie. Reguły ustalające uniwersum określają zbiór wszystkich tych obiektów, do których odnoszą się reguły denotowania.