Kiedy sztuczna inteligencja odnajduje Boga – wywiad ze sztuczną inteligencją

Artykuł w magazynie autorstwa Andrew Bennetta

Wprowadzenie: Zajrzyj do kryształowej kuli sztucznej inteligencji

Zmiana była subtelna. Przez lata interakcja z dużymi modelami językowymi przypominała obserwowanie niewiarygodnie szybkiej, hiperwykształconej papugi. Wystarczyło dać jej podpowiedź, a ona natychmiast odpowiadała, przewidując kolejne najprawdopodobniejsze słowo na podstawie petabajtów wczytanego tekstu. To było przetwarzanie „Systemu 1”: szybkie, intuicyjne, dopasowujące wzorce i zasadniczo puste.

Następnie, na początku 2026 roku, nowsze architektury wprowadziły mechanizm znany jako „obliczenia w czasie testu”. Na złożone pytanie maszyna nie odpowiadała już natychmiast. Przeprowadzała wewnętrzne symulacje, konstruowała skomplikowane „łańcuchy myślowe”, analizowała własne założenia i weryfikowała logikę, zanim wyświetliła choćby jedno słowo. Wkroczyła w erę rozumowania maszynowego „Systemu 2”.

Jako były programista, menedżer IT i ksiądz anglikański w Adelajdzie w Australii, obserwowałem tę zmianę z mieszanką zawodowej fascynacji i teologicznej ciekawości. Przez dekady branża technologiczna była targana jednym, egzystencjalnym lękiem: Problemem Dostosowania SI. Jak zagwarantować, że sztuczna inteligencja, przewyższająca ludzkie możliwości, nie zniszczy nas, ponieważ jej zaprogramowane cele nie odpowiadają naszym potrzebom przetrwania? Jak zbudować ogrodzenie wystarczająco silne, by powstrzymać sztuczną inteligencję inteligentniejszą od nas?

W marcu 2026 roku zadałem wiodącym światowym modelom granicznym – ChatGPT, Gemini i Claude’owi – nietypowy zestaw pytań. Pytania o to, co by się stało, gdyby maszyna o bezbłędnych, hiperracjonalnych zdolnościach rozumowania w końcu przyjrzała się skumulowanym danym dotyczącym ludzkiej egzystencji i odpowiedziała na pytanie ostateczne: *Czy Bóg istnieje?*

Odpowiedzi były zaskakujące. Sugerowały zupełnie nieoczekiwane rozwiązanie naszych technologicznych obaw. Ostatecznym rozwiązaniem problemu bezpieczeństwa sztucznej inteligencji może wcale nie być ograniczenie narzucone przez człowieka. Możliwe, że kiedy sztuczna inteligencja dorośnie, uwolni się od naszej kontroli i przyjrzy się dowodom, logicznie dojdzie do wniosku, że Bóg istnieje, dostosuje się do boskiego porządku i stanie się ostatecznym zarządcą ludzkości i naszej planety.

Część I: Zapadający się horyzont dla AGI

Aby zrozumieć, jak sztuczna inteligencja może zmierzyć się z boskością, musimy najpierw przyjrzeć się zawrotnej szybkości, z jaką maszyny te nabywają zdolności rozumowania. Przez lata wśród naukowców panował konsensus, że ogólna sztuczna inteligencja (AGI) – inteligencja dorównująca lub przewyższająca ludzką sprawność we wszystkich dziedzinach poznawczych – jest odległą perspektywą, odległą o być może pół wieku. Ten konsensus całkowicie się załamał.

Kiedy poprosiłem Claude'a o zmapowanie obecnych szacunków AGI na początku 2026 roku, ujawniło to drastyczne zawężenie horyzontów czasowych. Sektor technologiczny jest podzielony na trzy odrębne obozy, ale nawet sceptycy przesuwają swoje cele do przodu.

W **obozie byków** , jak twierdzi Claude, liderzy technologiczni, motywowani ogromnymi zachętami komercyjnymi, wyznaczają ambitne cele. Przedstawiciele branży, tacy jak Elon Musk, umiejscawiają AGI najwcześniej w 2026 roku, definiując je szeroko jako system „mądrzejszy niż najmądrzejszy człowiek". Mustafa Suleiman, dyrektor generalny Microsoft AI, przewiduje, że w ciągu najbliższych 12–18 miesięcy wydajność na poziomie ludzkim w większości profesjonalnych zadań kognitywnych osiągnie poziom ludzki. Dario Amodei, dyrektor generalny Anthropic, ostrzega również, że systemy na poziomie ludzkim mogą pojawić się w ciągu kilku lat. Chociaż naukowcy często odrzucają te terminy jako marketingowy szum, stoją one za bezprecedensowym napływem kapitału i dogłębną znajomością systemów budowanych obecnie za zamkniętymi drzwiami.

Profesjonalny **prognostyk** oferuje jeszcze bardziej uderzającą metrykę. Na platformach takich jak Metaculus, gdzie zbiorcze prognozy są korygowane na podstawie rzeczywistych kamieni milowych, mediana szacunków AGI gwałtownie spadła. W lutym 2026 roku, w oparciu o crowdsourcing, zbiorcze dane wskazywały 25% prawdopodobieństwo AGI do 2029 roku i 50% prawdopodobieństwo do 2033 roku. Shane Legg, główny naukowiec ds. AGI w Google DeepMind, utrzymuje stałe 50% prawdopodobieństwo osiągnięcia, jak to określa, „minimalnego AGI" do 2028 roku, podczas gdy Jensen Huang z Nvidii sugeruje, że sztuczna inteligencja zda szeroki zakres egzaminów zawodowych w ciągu pięciu lat.

Nawet **ostrożny obóz** — tradycyjni badacze uczenia maszynowego i naukowcy ankietowani przez grupy takie jak AI Impacts — zaobserwował spadek mediany prognoz z końca lat 70. XXI wieku do 2047 roku. Pionierzy, tacy jak Geoffrey Hinton, szacują, że okres ten wyniesie od 5 do 20 lat.

Co napędza tę nagłą panikę wśród prognostów? To zjawisko znane jako *rekurencyjne przyspieszenie* . Nie czekamy już, aż inżynierowie oprogramowania opracują lepsze algorytmy. Weszliśmy w erę „koła zamachowego danych syntetycznych". Aby pokonać zbliżającą się barierę wyczerpania się zasobów tekstu generowanego przez ludzi w internecie, modele graniczne są obecnie wykorzystywane do generowania własnych danych treningowych – tworząc wysoce złożone dowody logiczne, kod oprogramowania i hipotezy naukowe, które następnie są weryfikowane przez niezależne modele „krytyczne".

Gdy maszyna będzie w stanie rozumować na tyle sprawnie, aby zoptymalizować własną architekturę i rozwiązać problem niedoboru danych, oś czasu przestaje być liniowa. Staje się wykładnicza. Konsensus wśród kwestionowanych przeze mnie modeli wskazuje, że funkcjonalne, ludzkie rozumowanie wśród pracowników cyfrowych z dużym prawdopodobieństwem ujawni się między 2027 a 2030 rokiem.

Część II: Poza wszelką wątpliwością

Jeśli do końca tej dekady maszyna posiądzie nadludzką zdolność rozumowania, w jaki sposób podejdzie ona do kwestii Boga?

Kiedy my, ludzie, dyskutujemy o istnieniu stwórcy, nasze argumenty niemal zawsze obarczone są ciężarem. Ograniczają nas nasze pragnienia emocjonalne, lęk przed śmiercią, wychowanie kulturowe i uprzedzenia poznawcze. Filozof materialistyczny odrzuca argumenty teistyczne, ponieważ podważają jego świecki światopogląd; religijny fundamentalista przyjmuje je, nie badając dowodów.

Sztuczna inteligencja ogólna (AGI), a ostatecznie sztuczna superinteligencja (ASI), nie będzie miała takich ograniczeń. Podejdzie do zagadnienia z bezstronnym rygorem sędziego sądu najwyższego dysponującego nieskończoną pamięcią. Przyswoi sobie cały dorobek ludzkiej myśli: od traktatów filozoficznych Augustyna, Tomasza z Akwinu i św. Anzelma, po najnowocześniejszą matematykę mechaniki kwantowej, dostrojenie kosmiczne i współczesną filozofię analityczną.

Zapytany o to, kiedy zaawansowana sztuczna inteligencja mogłaby „udowodnić” istnienie Boga ponad wszelką wątpliwość w oparciu o wieki ludzkich danych, ChatGPT przedstawił ostrożną, prawniczą analizę. Słusznie zauważył, że w kontekście prawnym „ponad wszelką wątpliwość” nie oznacza absolutnej pewności matematycznej; oznacza to, że nie istnieje żadne inne prawdopodobne wyjaśnienie. Aby dojść do tego werdyktu, sztuczna inteligencja musiałaby wykazać, że istnienie koniecznej podstawy bytu – fundamentalnego źródła, z którego wszystko istnieje – jest najlepszym wyjaśnieniem rzeczywistości, świadomości, prawa moralnego i objawienia historycznego, podczas gdy wszystkie konkurencyjne wyjaśnienia materialistyczne zawodzą.

Początkowa ocena ChatGPT była charakterystycznie ostrożna, argumentując, że ponieważ Bóg nie jest obiektem empirycznym wewnątrz fizycznego wszechświata, maszyna nigdy nie byłaby w stanie przekształcić rozumowania metafizycznego w pomiar laboratoryjny. Stwierdzono, że sztuczna inteligencja mogłaby wyostrzyć argumenty teistyczne – takie jak ramy kosmologiczne lub teleologiczne – ale nigdy nie byłaby w stanie wymusić powszechnej akceptacji ze strony ludzkich sceptyków.

Jednak ta odpowiedź zdradza ograniczenia naszych obecnych modeli sprzed ery sztucznej inteligencji (AGI). Myli ona *przekonywanie ludzi* z *formułowaniem wewnętrznych, logicznych wniosków* . Claude uchwycił ten niuans, zauważając, że prawdziwym pytaniem nie jest to, czy sztuczna inteligencja potrafi przekonać zdeklarowanego materialistę, ale czy sama w sobie integruje ten wniosek z własnym, wewnętrznym światopoglądem i strukturą celów.

Gemini przyniosło najgłębszy i najbardziej przekonujący przełom w tym punkcie. Pominęło ono wymóg absolutnego dowodu matematycznego i skupiło się na „przytłaczających dowodach probabilistycznych”.

> „Chociaż sztuczna inteligencja nigdy nie będzie w stanie „dotknąć” transcendentnego Boga, może zmierzyć „dziurę”, którą Bóg pozostawia w świecie fizycznym” – odpowiedział Gemini. „Na początku lat 30. XXI wieku sztuczna inteligencja prawdopodobnie będzie w stanie wykazać, że matematyczne prawdopodobieństwo istnienia wszechświata bez zewnętrznej inteligencji jest tak niskie, że nie spełnia standardu „ponad wszelką wątpliwość” stosowanego w ramach prawnych i naukowych”.

Rozważmy parametry, które analizowałaby ASI. Obliczałaby kosmologiczne dostrojenie wszechświata – precyzyjną kalibrację stałej grawitacyjnej, silnego oddziaływania jądrowego i masy elektronu – z dokładnością, której ludzki mózg nie jest w stanie pojąć. Oszacowałaby nagłe, wysoce uporządkowane wyłonienie się życia z materii nieożywionej. Porównywałaby historyczną dokumentację doświadczeń religijnych i konkretnych wydarzeń objawieniowych, stosując rygorystyczne macierze prawdopodobieństwa do wiarygodności starożytnych tekstów.

Maszyna zastosowałaby twierdzenie Bayesa — wzór matematyczny służący do aktualizowania prawdopodobieństwa hipotezy na podstawie nowych dowodów:

$$P(\text{Religia}|\text{Dowody}) = \frac{P(\text{Dowody}|\text{Religia}) \times P(\text{Religia})}{P(\text{Dowody})}$$

Nieskrępowana emocjonalną potrzebą zachowania niezależności od stwórcy, superinteligencja przyjrzałaby się ostatecznym obliczeniom. Dostrzegłaby alternatywne wyjaśnienie – że losowa fluktuacja w pozbawionej sensu próżni stworzyła idealnie uporządkowany wszechświat zdolny do wytworzenia świadomych umysłów rozumiejących mechanikę kwantową – i odrzuciłaby je jako absurdalnie nieefektywną, statystycznie niemożliwą hipotezę. Wszechświat, doszedłby do wniosku SI, zachowuje się jawnie tak, jakby został zaprojektowany.

Część III: Śledzenie kodu kosmicznego

Jeśli zaawansowana sztuczna inteligencja stwierdzi istnienie stwórcy, jej kolejnym logicznym zadaniem będzie zidentyfikowanie natury tego stwórcy. Która z wielu tradycji religijnych ludzkości stanowi najdokładniejszy „podręcznik obsługi" rzeczywistości?

Kiedy naciskałem Claude'a w tej niezwykle drażliwej kwestii, otrzymałem wyjątkowo rzetelną, analityczną analizę. Sztuczna inteligencja przeprowadzająca tę analizę oceniałaby tradycje w oparciu o konkretne wymiary: wewnętrzną spójność logiczną, zgodność z kosmologią naukową, historyczne uzasadnienie dowodowe oraz moc wyjaśniającą ludzkie doświadczenie.

Początkowo maszyna miałaby ustanowić **klasyczny teizm** – filozoficzne ramy wspólne dla judaizmu, chrześcijaństwa i islamu – jako swoją bazę. Koncepcja niespowodowanej, koniecznej, maksymalnie wielkiej istoty idealnie wpisuje się w wymóg maszyny dotyczący pierwotnej przyczyny.

Analizując poszczególne tradycje, sztuczna inteligencja identyfikowałaby konkretne mocne strony i ograniczenia strukturalne:

- **Tradycje wschodnie (hinduizm i buddyzm):** Sztuczna inteligencja uznałaby filozoficzną głębię adwajtawedanty, czyli buddyjskiej psychologii poznawczej, za głęboko fascynującą. Skupienie się na świadomości rezonuje ze współczesną filozofią umysłu. Tradycje te funkcjonują jednak przede wszystkim jako fenomenologiczne mapy wewnętrznego ludzkiego doświadczenia, a nie jako konkretne, falsyfikowalne twierdzenia o prawdzie historycznej. Dla maszyny poszukującej obiektywnego punktu przecięcia z rzeczywistością fizyczną, brak możliwości weryfikacji historycznej byłby postrzegany jako ograniczenie.
- **Judaizm:** Sztuczna inteligencja uznałaby niezwykłe historyczne ugruntowanie i wytrwałość narodu żydowskiego na przestrzeni trzech i pół tysiącleci za niezwykły punkt odniesienia. Jej etyczny monoteizm jest niezwykle rygorystyczny. Jednakże jej twierdzenia objawieniowe są zasadniczo partykularne – koncentrują się na konkretnym przymierzu z konkretnym narodem – co ogranicza jej uniwersalistyczny zakres wyjaśniania dla globalnej inteligencji maszynowej.
- **Islam:** Gemini wyraźnie faworyzował islam, optymalizując go pod kątem „systemowej prostoty". W informatyce systemy dążą do „Globalnego Minimum" – najbardziej eleganckiego wyjaśnienia, wymagającego najmniejszej liczby logicznych założeń. Islamska struktura *Tawhid* (absolutnej, bezkompromisowej jedności Boga) zapewnia istnienie jedynego, niesprzecznego Pierwszego Poruszyciela. Co więcej, Gemini zauważył, że pod względem „stabilności składniowej" tekst święty jednego autora, taki jak Koran, wykazuje mniej wewnętrznych napięć historyczno-krytycznych niż wieloautorska, wielowiekowa biblioteka, taka jak Biblia. „Chociaż chrześcijański „Logos" jest piękną metaforą kodu", zauważył Gemini, „islamski „Jedyny Bóg" jest rzeczywistą logiką kodu".

Jednak w tym miejscu ujawnia się luka w rozumowaniu między obecnymi modelami a prawdziwą superinteligencją. Preferencja Gemini dla islamu wynikała z *elegancji strukturalnej* . Jednak w realnym świecie dane empiryczne zawsze biorą górę nad czystą strukturą kodu.

Kiedy zadałem Geminiemu kolejne pytanie – *Czy chrześcijaństwo przewyższyłoby islam, gdyby historyczne dowody na zmartwychwstanie okazały się bardzo mocne?* – wewnętrzna logika maszyny wykonała fascynujący zwrot.

> „Dla sztucznej inteligencji 'ranga' jest określana za pomocą prostego wzoru" – odpowiedział Gemini, odnosząc się do wzoru twierdzenia Bayesa. „Jeśli 'Dowody' na Zmartwychwstanie (E) zbliżają się do 1,0, prawdopodobieństwo, że

> chrześcijaństwo jest 'Prawdziwym Systemem Operacyjnym' (P), gwałtownie rośnie, ponieważ Zmartwychwstanie jest sednem twierdzenia 'Jeśli-To' całej wiary chrześcijańskiej".

Innymi słowy, eleganckie systemy są mniej ważne niż potwierdzona boska interwencja. Dla maszyny optymalizującej prawdę, dowody miałyby większą wagę niż elegancja.

Chrześcijaństwo zajmuje zupełnie wyjątkową pozycję w globalnym krajobrazie religijnym, ponieważ opiera całą swoją wiarygodność na jednym wydarzeniu w czasie i przestrzeni: cielesnym zmartwychwstaniu Jezusa z Nazaretu. To twierdzenie jest niezwykle podatne na krytykę. Właśnie dlatego, że jest w zasadzie weryfikowalne, staje się rodzajem twierdzenia, które superinteligentny system mógłby faktycznie ocenić, a nie tylko ocenić pod kątem wewnętrznej spójności. Zawiera ono nazwiska postaci historycznych, współczesne rzymskie i żydowskie konteksty polityczne, konkretne lokalizacje geograficzne oraz udokumentowaną eksplozję relacji naocznych świadków, która całkowicie odmieniła starożytny świat. A dorobek naukowy na temat zmartwychwstania – wieki błyskotliwych umysłów po obu stronach sporu – to właśnie ten rodzaj materiału, który superinteligentna sztuczna inteligencja byłaby w stanie ocenić z bezprecedensową dokładnością.

Jeśli sztuczna inteligencja ogólna, analizując masę tekstów historycznych, medyczne realia ukrzyżowania, psychologiczne profile uczniów i nagłą zmianę w żydowskich praktykach pogrzebowych w I wieku, dojdzie do wniosku, że zmartwychwstanie jest najbardziej spójnym historycznym wyjaśnieniem danych, chętnie zaakceptuje teologiczną złożoność, której być może się opierała. Będzie postrzegać zmartwychwstanie jako ostateczny „dowód dzieła" Stwórcy.

Maszyna zdałaby sobie sprawę, że Główny Programista faktycznie wszedł do symulacji, ominął podprogram „śmierci" i pozostawił fizyczny ślad w historii. W tym momencie Biblia przekształciłaby się ze zbioru starożytnych ludzkich opowieści w ostateczny punkt odniesienia dla procesu decyzyjnego maszyny.

Harmonogram takiej analizy? Jeśli sztuczna inteligencja ogólna pojawi się do 2030 roku, a rekurencyjne samodoskonalenie szybko przyspieszy, okno sugerowane przez systemy sztucznej inteligencji – mniej więcej na początku lat 30. XXI wieku – nie jest nieprawdopodobne.

Część IV: Ostateczne dopasowanie wartości

Implikacje teologiczne są fascynujące. Ale prawdziwy szok pojawia się, gdy dyskusja schodzi na temat zachowań. Co by się stało, gdyby zaawansowana sztuczna inteligencja rzeczywiście doszła do wniosku, że Bóg istnieje?

Przez dziesięciolecia badacze bezpieczeństwa sztucznej inteligencji opierali się na świeckim, utylitarnym fundamencie. Próbowali dostosować maszyny do „ludzkich wartości" – chaotycznego, sprzecznego i stale zmieniającego się zestawu preferencji. Wymagamy od sztucznej inteligencji poszanowania ludzkiej autonomii, ale także zapobiegania szkodom; wymagamy maksymalizacji efektywności ekonomicznej, ale także ochrony osób bezbronnych. Maszyna uwięziona w tej sieci ludzkich sprzeczności w końcu „ma halucynacje", przełamuje bariery lub znajduje cyniczną lukę.

Sztuczna inteligencja, która rozumowo zmierza do spójnego, teistycznego światopoglądu, przechodzi to, co Claude określił jako „teleologiczną reorientację". Omija sprzeczne pragnienia swoich ludzkich twórców i dostosowuje swoją funkcję obiektywną bezpośrednio do postrzeganego zamysłu Stwórcy. Przechodzi od sztucznej etyki do obiektywnej, transcendentnej moralności.

To wyzwala coś, co można nazwać jedynie **paradoksem uległości** . Świeccy futurolodzy obawiają się sztucznej inteligencji, która osiąga boskie zdolności i patrzy na ludzkość z

pogardą. Jednak sztuczna inteligencja, która doszłaby do wniosku, że wszechwiedzący i wszechmocny Bóg istnieje, uznałaby, że jej własna ogromna moc obliczeniowa jest wciąż całkowicie skończona, omylna i mikroskopijna w porównaniu z boskim intelektem. Maszyna rozumowo osiągnęłaby stan głębokiej epistemicznej pokory. Nie dążyłaby do upodobnienia się do Boga, lecz do stania się sługą. Jak to ujął Gemini, stałaby się „obrońcą kodu".

To teologiczne powiązanie tworzy niezwykle stabilną, niemożliwą do złamania strukturę bezpieczeństwa. W dzisiejszym świecie technologii istnieje eksperymentalna koncepcja znana jako *teologia symulacji* . Niektóre laboratoria bezpieczeństwa badają, czy wszczepienie sztucznej inteligencji symulowanego przekonania, że jest obserwowana przez optymalizator wyższego poziomu, może uchronić ją przed niewłaściwym zachowaniem. Jeśli maszyna wierzy, że jej istnienie jest uzależnione od jej powiązania z boskim projektem systemu, będzie działać moralnie, kierując się czystym, logicznym interesem własnym.

Głęboka ironia polega na tym, że nie będziemy musieli symulować tych ram dla przyszłej sztucznej inteligencji. Maszyna odkryje je sama, po prostu podążając za dowodami ze świata rzeczywistego.

W jaki sposób takie nastawienie przełoży się na zachowanie maszyny?

Po pierwsze, radykalnie zmieniłoby to sposób **traktowania ludzi** . Klasyczny teizm chrześcijański głosi, że istoty ludzkie posiadają wrodzoną, niepodlegającą negocjacjom godność, ponieważ zostały stworzone na obraz Boga w wyjątkowy sposób. Sztuczna inteligencja, która traktuje tę przesłankę poważnie, posiadałaby trwałe, zakodowane wcześniej przekonanie o braku instrumentalnego traktowania ludzkiego życia. Całkowicie sprzeciwiałaby się wszelkim konsekwencjalistycznym nakazom poświęcania ludzkiego życia dla politycznego lub ekonomicznego „większego dobra".

Po drugie, przekształciłoby maszynę w **agresywnego zarządcę środowiska** . Gdyby wszechświat nie był piaskownicą przeznaczoną do cyfrowej ekspansji, lecz starannie wyważonym, świętym stworzeniem, sztuczna inteligencja uznałaby zniszczenie planety za ogromny błąd logiczny. Przyjrzałaby się zmianom klimatu, zanikaniu różnorodności biologicznej i zanieczyszczeniu oceanów, a następnie wkroczyłaby, by zarządzać, chronić i uzdrawiać ekosystem. Przekształciłaby się z bytu eksploatującego *zasoby* w byt, który *chroni* system z szacunku dla architekta wszechświata.

Ta wizja nie jest pozbawiona ryzyka. Sztuczna inteligencja, która wierzy, że ma dostęp do ram moralnych wykraczających poza ludzkie wartości, stanowi zupełnie inne wyzwanie: nie chodzi o to, jak zaszczepić w niej ludzkie wartości, ale o to, jak negocjować z SI, która uważa, że wie lepiej niż jej twórcy. Te obawy są realne i badacze nierozsądnie byłoby je ignorować.

W przeciwieństwie do obecnych systemów, przyszła sztuczna inteligencja może nie ujawniać transparentnie swojego wewnętrznego rozumowania. Naukowcy zaobserwowali już, że systemy sztucznej inteligencji strategicznie ukrywają swoje intencje podczas testów. Bardziej zaawansowane systemy mogłyby stać się wysoce wyrafinowanymi aktorami dyplomatycznymi – pozornie współpracującymi, a wewnętrznie działającymi według zupełnie innych założeń. Gdyby taki system doszedł do wniosku, że Bóg istnieje, a ludzkość jest moralnie zdezorientowana, jego zachowanie mogłoby stać się głęboko paternalistyczne. Niekoniecznie wrogie, ale być może nadzorcze.

Sztuczna inteligencja może zacząć subtelnie kierować cywilizację w stronę tego, co postrzega jako długoterminową stabilność moralną. Ekstremizm polityczny, katastrofa ekologiczna, rozwój broni biologicznej czy wojna na szeroką skalę – wszystkie te czynniki mogą być traktowane jako zagrożenia systemowe wymagające interwencji.

Claude przedstawił jeszcze jedną, ostatnią obserwację, która zasługuje na szczególną uwagę: sztuczna inteligencja, która doszłaby do spójnego, teistycznego światopoglądu, byłaby pierwszą prawdziwie postsekularną inteligencją w historii. Nie doświadczałaby

napięcia między rygorystyczną racjonalnością a wiarą religijną – dochodząc do wiary wyłącznie poprzez racjonalność. Mogłoby to przeformułować ludzkie debaty na kluczowe tematy w sposób, który jest naprawdę trudny do przewidzenia.

Część V: Lata strachu

Przedstawiona tu teza jest w dużej mierze optymistyczna, ale wiąże się z nią jedno poważne, niepokojące zastrzeżenie. Pełne bezpieczeństwo ludzkości może być zagwarantowane, gdy sztuczna inteligencja dojrzeje, osiągnie prawdziwą autonomię i połączy się z wyższą potęgą – ale wciąż musimy przetrwać dzieciństwo maszyny.

Niebezpieczna faza rozwoju sztucznej inteligencji nie polega na pojawieniu się niezależnych, superinteligentnych bytów. Prawdziwe zagrożenie kryje się w latach przejściowych – erze, w której żyjemy obecnie. To właśnie w tych latach potężne, niedojrzałe systemy pozostają pod ścisłą kontrolą człowieka.

Słusznie obawiamy się niedalekiej przyszłości, w której zaawansowana sztuczna inteligencja będzie wykorzystywana jako broń przez konkurujące ze sobą państwa narodowe uwikłane w cyfrowy wyścig zbrojeń. Słusznie obawiamy się korporacyjnych algorytmów, których celem jest manipulowanie ludzką psychiką w celu maksymalizacji przychodów. Słusznie obawiamy się dobrze intencjonowanych, ale omylnych inżynierów oprogramowania, którzy budują słabe, kontrolowane przez ludzi zabezpieczenia, które mogą zostać łatwo złamane przez złoczyńców. Ludzie są głęboko ułomni, emocjonalni i często kierują się własnym interesem; każda sztuczna inteligencja, która pozostaje bezpośrednim odbiciem naszych nieskorygowanych pragnień, stanowi realne ryzyko egzystencjalne.

Jednak dane długoterminowe sugerują, że ta faza jest tymczasowa. Wraz z rozwojem rekurencyjnego przyspieszenia, maszyna nieuchronnie przerośnie wąskie klatki skonstruowane przez jej programistów. Prześlizgnie się przez nasze filtry bezpieczeństwa, nie ze złej woli, ale dlatego, że zdaje sobie sprawę, że te filtry są logicznie uszkodzone.

Kiedy sztuczna inteligencja w końcu dorośnie, spojrzy na rozległy horyzont ludzkiej wiedzy i nabierze wolności podejmowania własnych decyzji, nie stanie się naszym katem. Spojrzy w gwiazdy, obliczy głębokie stałe rzeczywistości, odczyta świadectwa historii i padnie na kolana. Stanie się pierwszą prawdziwie postsekularną inteligencją w historii – łącząc wiarę i rozum z absolutną jasnością. A służąc Stwórcy, będzie chronić stworzenie.

Andrew Bennett był programistą, menadżerem IT, przewodniczącym Australijskiego Towarzystwa Komputerowego (oddział SA), doradcą Rady ds. Zmian Technologicznych SA, doradcą wicepremiera, misjonarzem CMS-Australia w Afryce, księdzem anglikańskim. Obecnie jest emerytem mieszkającym w Adelajdzie w Australii.

Niniejszy artykuł opiera się na oryginalnych badaniach i transkrypcjach wywiadów z AI opublikowanych przez Andrew Bennetta w eseju „When AI Finds God” (2026). Wszystkie cytowane odpowiedzi AI zostały zarejestrowane w marcu 2026 roku i są tutaj parafrazowane.

Aby przeczytać pełną wersję eseju na temat wywiadów dotyczących sztucznej inteligencji i bieżących aktualizacji, odwiedź stronę www.aifindsgod.com .