Kod jako norma: algorytmiczne rządzenie w świetle myśli Sandela i tradycji chińskiej



AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje przejście od tradycyjnych norm prawnych i społecznych do architektury technicznej, gdzie kod staje się formą sterowania zachowaniem ex ante. Autor argumentuje, że algorytmy nie są neutralne, lecz realizują konkretne koncepcje dobra zakodowane w ich funkcjach celu. Wykorzystując myśl Michaela Sandela oraz chińskie pojęcia takie jak Jilin (mechaniczne serce-umysł), tekst wskazuje na ryzyko instrumentalizacji człowieka i utraty wewnętrznego dobra praktyk zawodowych. Główną tezą jest stwierdzenie, że technokracja nie zastępuje konfliktów wartości obliczeniami, lecz ukrywa filozofię polityczną w specyfikacji systemów. Wobec tego niezbędna jest teleologiczna analiza AI, aby zapobiec moralnej depersonalizacji i nadmiernej optymalizacji kosztem godności oraz relacji międzyludzkich.

The article analyzes the transition from traditional legal and social norms to technical architecture, where code becomes a form of ex ante behavioral control. The author argues that algorithms are not neutral but rather implement specific conceptions of the good encoded in their objective functions. Drawing on the thought of Michael Sandel and Chinese concepts such as Jilin (mechanical heart-mind), the text points to the risk of human instrumentalization and the loss of the internal goods of professional practices. The main thesis is that technocracy does not replace value conflicts with calculations, but rather hides political philosophy within system specifications. Consequently, a teleological analysis of AI is essential to prevent moral depersonalization and excessive optimization at the expense of dignity and interpersonal relationships.

Artykuł analizuje transformację kontroli społecznej w erze cyfrowej, gdzie tradycyjne mechanizmy normatywne – rytuał (li) i prawo (fa) – zostają uzupełnione przez 'kod', czyli architekturę techniczną systemów AI. Autor stawia tezę, że algorytmiczne zarządzanie nie eliminuje polityki ani moralności, lecz maskuje je w technicznych specyfikacjach i funkcjach celu, tworząc złudzenie obiektywizmu. Poprzez syntezę myśli konfucjańskiej, daoistycznej oraz współczesnej filozofii politycznej (m.in. M. Sandela), tekst dowodzi, że predykcja danych nie jest tożsama z rozumieniem człowieka jako podmiotu moralnego. Systemy AI, optymalizując konkretne wskaźniki, mogą prowadzić do 'mechanizacji serca-umysłu' (jixin) i redukcji osoby do profilu probabilistycznego, co zagraża autonomii jednostki oraz wspólnej przestrzeni deliberatywnej. W obliczu skalowalności błędów algorytmicznych, autor postuluje przywrócenie teleologicznej interpretacji praktyk społecznych oraz uznanie, że odpowiedzialność moralna nie może być delegowana na systemy, a mądrość współczesnego ładu powinna objawiać się w zdolności do wyznaczania granic optymalizacji.

SŁOWA KLUCZOWE

algorytmiczne rządzenie | kod jako prawo | telos praktyki | jixin | antropologia cyfrowa | predykcja a rozumienie | wyjaśnialność techniczna | contestability | personalizacja algorytmiczna | skalowalność normatywnego błędu | epistemiczne li | relacyjność moralna | governance AI | architektura widzialności | moralna depersonalizacja

Kod jako nowa forma normatywnej kontroli

Gdy norma staje się kodem: sztuczna inteligencja, algorytmiczne rządzenie i nowa postać jixin

Dotychczasowy spór między li i fa zakładał, że normę społeczną przekazuje człowiek lub instytucja. Li działało poprzez wychowanie, rytuał, nawyk, wstyd i relację; fa zaś przez jawny standard, nakaz, zakaz i sankcję. Cyfrowa nowoczesność wprowadza jednak trzeci mechanizm normatywny: architekturę techniczną, która nie tylko wskazuje człowiekowi właściwy sposób działania, lecz częściowo organizuje przestrzeń jego możliwości, zanim podejmie on świadomy wybór. Algorytm decydujący o kolejności informacji, przydziale zlecenia, widoczności treści, sugerowanej trasie czy

prawdopodobieństwie uzyskania oferty nie zachowuje się jak klasyczny prawodawca. Nie wydaje ogólnej normy wszystkim jednakowo. Zarządza środowiskiem decyzji.

Właśnie dlatego Lawrence Lessig mógł pisać o "kodzie jako prawie", a współczesna teoria regulacji zaczęła traktować architekturę informatyczną jako odrębny sposób sterowania zachowaniem. Prawo mówi: nie rób. Norma społeczna mówi: nie wypada. Rynek mówi: to będzie kosztować. Kod natomiast może sprawić, że określona możliwość stanie się trudniej dostępna, mniej widoczna lub zostanie całkowicie wyłączona. Norma prawna tradycyjnie zakłada podmiot, który może ją naruszyć i ponieść konsekwencje; architektura cyfrowa działa coraz częściej ex ante: przebudowuje pole wyboru, zanim naruszenie w ogóle stanie się możliwe.

Nie oznacza to, że algorytm jest nowym prawodawcą w dosłownym znaczeniu. Algorytm sam w sobie nie posiada celu moralnego; realizuje funkcję określoną przez projektantów, przedsiębiorstwo, urząd lub konfigurację procesu uczenia. Dlatego kluczowe pytanie nie brzmi: "czy algorytm jest sprawiedliwy?", lecz: jaką koncepcję dobra zakodowano w funkcji, którą system ma optymalizować, jakie dane uznano za istotne, jakie błędy za dopuszczalne i czyje interesy potraktowano jako koszt uboczny. Problem algorytmiczny okazuje się zatem głęboko Sandelowski. Nie można rozstrzygnąć, jak powinien działać system, dopóki nie odpowiemy, jaki jest telos praktyki, którą ten system organizuje.

System rekomendacyjny projektowany dla maksymalizacji zaangażowania posiada inny telos niż system służący jakości informacji. Algorytm dostaw optymalizujący czas realizacji zamówienia może traktować oczekiwanie klienta jako koszt, lecz bezpieczeństwo kuriera również jest dobrem. Model kredytowy minimalizujący straty instytucji finansowej odpowiada na inne pytanie niż instytucja publiczna zainteresowana dostępnością kapitału dla grup historycznie niedofinansowanych. Matematyczna funkcja celu nie usuwa normatywności; koncentruje ją w parametrze, czyniąc moralny wybór mniej widocznym, gdyż zostaje on zapisany w języku technicznym.

Tu ujawnia się jedno z najważniejszych złudzeń technokracji: przekonanie, że model może zastąpić konflikt wartości trafniejszym obliczeniem. Tymczasem obliczenie zaczyna się dopiero po określeniu tego, co ma być maksymalizowane, minimalizowane lub przewidywane. Sztuczna inteligencja może pomagać w znalezieniu środka do celu, lecz nie wyprowadzi z danych samego obowiązku moralnego, który cel należy uznać za właściwy. Nie istnieje zbiór danych, z którego logicznie wynikałoby, ile bezpieczeństwa należy poświęcić na rzecz szybkości, ile prywatności na rzecz wygody ani jaki stopień nierówności jest społecznie dopuszczalny. Algorytmiczna polityka nie eliminuje filozofii politycznej; ukrywa ją w specyfikacji systemu.

W tym punkcie konfucjańska teleologia roli spotyka Sandela. Dobry lekarz, nauczyciel, urzędnik czy rodzic nie może zostać oceniony wyłącznie według wydajności dowolnego wskaźnika, ponieważ rola posiada wewnętrzne dobro. Analogicznie system używany w edukacji nie powinien być uznawany za dobry tylko dlatego, że zwiększa przewidywalność wyniku egzaminacyjnego; trzeba zapytać, czy jego działanie sprzyja uczeniu się. System wspomagający medycynę nie jest dobry tylko dlatego, że obniża koszt pojedynczej procedury; istotne pozostają zdrowie, trafność diagnozy, godność i relacja terapeutyczna. AI nie zwalnia więc z teleologicznej interpretacji praktyki, lecz czyni ją pilniejszą.

Daoistyczne jixin dostarcza temu problemowi szczególnie interesującej heurystyki. Książka definiuje jixin jako "mechaniczne serce-umysł": sposób myślenia skoncentrowany na efektywności, który zaczyna traktować naturę i ludzi instrumentalnie. W dalszej części notatki pojęcie to zostaje wykorzystane przy analizie biotechnologicznej optymalizacji człowieka oraz zestawione z ziran, zhi zu i zhi zhi – spontanicznością, umiarem i zdolnością rozpoznania momentu, w którym należy przestać optymalizować. Przeniesienie tej intuicji na sztuczną inteligencję jest interpretacyjnym rozszerzeniem, a nie twierdzeniem historycznym: Zhuangzi nie "przewidział AI". Jego krytyka pozwala jednak zapytać, co dzieje się z kulturą, gdy coraz więcej dziedzin zaczyna być postrzeganych jako problemy optymalizacyjne. Niebezpieczeństwo jixin nie polega bowiem na samym używaniu maszyny.

Kucharz Ding z Zhuangzi jest mistrzem techniki.

Profil algorytmiczny nie zastępuje podmiotowości moralnej człowieka

Problemem jest utrata zdolności do rozróżnienia sposobu pomiaru świata od samego świata. Jeśli system postrzega pracownika przez liczbę zrealizowanych zadań, ucznia przez wynik, klienta przez prawdopodobieństwo zakupu, obywatela przez klasyfikację ryzyka, a użytkownika platformy przez profil zainteresowań, każda z tych reprezentacji może być funkcjonalnie użyteczna. Błąd pojawia się jednak w momencie, gdy model osoby zostaje utożsamiony z samą osobą. Powstaje wówczas szczególna antropologia cyfrowa.

Tradycyjny konfucjanizm zakładał, że człowiek istnieje poprzez relacje. Platforma cyfrowa może sformułować tezę pozornie podobną: użytkownik jest sumą swoich kliknięć, kontaktów, historii wyszukiwania, transakcji, czasu zatrzymania wzroku i podobieństwa do innych osób. Oba modele są „relacyjne”, lecz ich natura jest zupełnie inna. Relacja konfucjańska ma charakter normatywny i interpretacyjny; profil algorytmiczny jest probabilistyczny. Ojciec, córka, przyjaciel czy obywatel to

role nasycone znaczeniem. Z kolei „użytkownik z 83-procentowym prawdopodobieństwem zainteresowany produktem X” to jedynie klasyfikacja operacyjna.

Nie wolno zatem utożsamiać relacyjności danych z relacyjnością moralną. Platforma może dysponować ogromną wiedzą o sieci połączeń człowieka, a mimo to nie rozumieć moralnego sensu żadnej z nich. Może wykryć częstotliwość kontaktów z matką, nie wiedząc, czy są one przejawem troski, konfliktu czy obowiązku. Może przewidzieć zmianę pracy, lecz nie pojąć, czy wynika ona z aspiracji, wypalenia czy odpowiedzialności rodzinnej.

Dane zachowują ślady relacji, ale nie wyczerpują ich sensu. To rozróżnienie prowadzi do kluczowej konstatacji: predykcja nie jest rozumieniem. System może prawidłowo przewidzieć zachowanie człowieka, nie posiadając adekwatnej teorii jego racji. Z perspektywy zarządzania różnica ta może wydawać się nieistotna – jeśli przewidywanie działa, cel zostaje osiągnięty. Z perspektywy moralnej jest ona zasadnicza, ponieważ człowiek nie chce być traktowany wyłącznie jako obiekt trafnej prognozy. Pragnie statusu podmiotu, którego racje mogą zostać wysłuchane.

Właśnie tutaj ujawnia się nieoczywiste znaczenie prawa do wyjaśnienia i zakwestionowania decyzji automatycznej. Chińska ustawa o ochronie danych osobowych nakazuje zachowanie przejrzystości oraz sprawiedliwości przy automatycznym podejmowaniu decyzji, zakazując jednocześnie nieuzasadnionego różnicowania warunków transakcji. Jeśli automatyczna decyzja istotnie wpływa na prawa i interesy osoby, może ona żądać wyjaśnień oraz sprzeciwić się rozstrzygnięciu podjętemu wyłącznie automatycznie. To interesująca instytucjonalizacja różnicy między osobą jako profilem a osobą jako uczestnikiem normatywnej relacji: klasyfikowany człowiek ma otrzymać możliwość ponownego wejścia do procesu jako ktoś zdolny przedstawić własne racje.

Wyjaśnialność techniczna nie zastępuje uzasadnienia moralnego

Samo wyjaśnienie techniczne nie rozwiązuje jednak problemu. Można precyzyjnie opisać model, a wciąż nie odpowiedzieć na pytanie, czy zastosowane kryterium jest moralnie właściwe. Stwierdzenie, że ktoś nie otrzymał określonej możliwości, ponieważ model przypisał mu ryzyko na poziomie 0,72, mówi niewiele – dopóki nie wiemy, dlaczego akurat taki próg uznano za słuszny, jaki jest koszt błędu fałszywie pozytywnego i kto ponosi jego konsekwencje.

Wyjaśnialność matematyczna i uzasadnialność moralna to dwie różne kwestie. Dlatego w normatywnym zarządzaniu AI ważniejsza od pełnej technicznej przejrzystości bywa niekiedy contestability, czyli możliwość skutecznego zakwestionowania wyniku. Człowiek nie musi znać

wszystkich parametrów modelu, podobnie jak pacjent nie musi rozumieć równań urządzenia diagnostycznego. Powinien natomiast wiedzieć, jakie kryteria dotyczą jego sytuacji, jaki jest skutek decyzji, kto za nią odpowiada, gdzie można zgłosić błąd i czy istnieje człowiek posiadający realną kompetencję do zmiany rezultatu. "Human in the loop" jest pustą formułą, jeżeli człowiek jedynie mechanicznie zatwierdza rekomendację systemu.

Chińskie regulacje algorytmiczne wyraźnie dostrzegają część tego problemu. Obowiązujące od marca 2022 roku przepisy dotyczące usług rekomendacji wymagają informowania użytkowników o korzystaniu z algorytmów, ujawnienia podstawowych zasad i mechanizmów ich działania oraz zaoferowania możliwości wyłączenia personalizacji na rzecz wariantu niezależnego od cech użytkownika. Umożliwiają one również wybór lub usunięcie przypisanych etykiet profilujących. Jest to konstrukcja interesująca filozoficznie: algorytm nie może pozostawać całkowicie niewidzialnym środowiskiem, skoro użytkownik ma zachować choćby część sprawczości wobec sposobu, w jaki jest profilowany.

Te same regulacje pokazują jednak, że chiński model zarządzania algorytmami wykracza poza liberalną ochronę autonomii. Obejmuje on jednocześnie prawa użytkowników, ochronę konsumentów, ograniczanie dyskryminacji i mechanizmów uzależniających, bezpieczeństwo oraz obowiązek respektowania interesu publicznego. Przepisy wymagają m.in. kontroli modeli skłaniających do nadmiernej konsumpcji i ustanawiają nadzór nad rekomendacją treści. Jest to architektura łącząca język praw podmiotowych z wyraźnie formacyjną koncepcją środowiska informacyjnego. W tym miejscu analogia do statecraft as soulcraft staje się szczególnie silna. Dawne państwo kształtowało obywatela poprzez rytuał, szkołę, prawo i przykład elity.

Dziś państwo oraz platformy cyfrowe oddziałują również poprzez architekturę widzialności: to, które informacje użytkownik zobaczy najpierw, co zostanie uznane za popularne i jak często system będzie wzmacniał określony typ zachowania. Formacja charakteru nie wymaga już jawnego wykładu moralnego; może odbywać się przez tysiące mikrodecyzji dotyczących kolejności treści i bodźców. W ten sposób technologia staje się narzędziem dyscyplinującym, co pozwala przywołać analizy Michela Foucaulta, który badał władzę poprzez organizację przestrzeni, czasu i obserwacji.

Personalizacja algorytmiczna jako ukryta forma władzy i odpowiedzialności

Algorytmiczne zarządzanie rozszerza tę logikę, gdyż nie musi stosować identycznego środowiska wobec wszystkich – pozwala na personalizację bodźców. Podczas gdy klasyczny regulamin jest

generalny, algorytm może być indywidualny. To jedna z fundamentalnych różnic między fa a „kodem". Fa w swoim ideałotypie dąży do jawnego standardu; system algorytmiczny potrafi natomiast dostosowywać regułę działania w czasie rzeczywistym, opierając się na historii konkretnego użytkownika.

Taka personalizacja rodzi paradoks sprawiedliwości. Z jednej strony może poprawiać adekwatność decyzji, gdyż ludzie różnią się potrzebami, a identyczne traktowanie nie zawsze jest sprawiedliwe. Z drugiej jednak im bardziej system personalizuje warunki, tym trudniej porównać własną sytuację z położeniem innych. Dyskryminacja staje się mniej widoczna, ponieważ znika jedna, publiczna reguła dostępna dla wszystkich. Właśnie dlatego chińska ustawa o ochronie danych osobowych oraz regulacje algorytmiczne wymagają sprawiedliwości w automatycznym podejmowaniu decyzji i zakazują nieuzasadnionego różnicowania cen oraz warunków wobec konsumentów. Tu ujawnia się problem, którego starożytna tradycja fa nie mogła znać: norma może zostać zindywidualizowana do tego stopnia, że przestaje być wspólnie obserwowalna. Publiczność prawa pozwala obywatelom wiedzieć nie tylko to, co wolno im, lecz również co wolno innym i administracji.

Algorytmiczna personalizacja potrafi stworzyć miliony wariantów środowiska normatywnego. Człowiek zna swój ekran, swoją cenę, ranking i ofertę, ale niekoniecznie regułę systemu jako całości, co osłabia społeczną zdolność rozpoznawania wzorców niesprawiedliwości. Problem ten jest szczególnie wyrazisty w pracy platformowej. Przepisy dotyczące rekomendacji algorytmicznych w Chinach wprost odnoszą się do pracowników, wymagając od dostawców usług odpowiednich mechanizmów w zakresie przydzielania zleceń, wynagrodzeń, godzin pracy oraz nagród i kar. Nie jest to przypadkowy szczegół regulacyjny.

Pokazuje on, że algorytm przejmuje funkcję niegdyś pełniącą przełożony: rozdziela zadania, ocenia wydajność, narzuca presję czasu i pośrednio kształtuje dochód. Badania empiryczne nad chińskim rynkiem pracy platformowej potwierdzają wagę tego zjawiska. Analiza dostawców żywności opisuje ścisłe terminy generowane przez systemy, długie godziny pracy oraz niepewność statusu, a także strategie pracowników polegające na „graniu" z algorytmem – np. poprzez manipulowanie lokalizacją lub wymienianie się zleceniami. Badanie jakościowe kurierów z 2025 roku wskazuje natomiast na głęboki rozdźwięk między logiką systemu a rzeczywistym doświadczeniem pracy. Międzynarodowa Organizacja Pracy zauważa z kolei, że algorytmy i wielowarstwowe pośrednictwo tworzą nowe formy podporządkowania, które nie mieszczą się w tradycyjnej binarnej kategorii pracownik-samozatrudniony.

Mamy tu do czynienia z głęboką zmianą w teorii władzy. Tradycyjny pracownik wiedział przynajmniej, kto wydaje polecenie. Pracownik platformowy doświadcza kontroli poprzez aplikację, której decyzja jest wypadkową szeregu modeli, danych, celów biznesowych i interwencji ludzkich.

Władza zostaje rozproszona, lecz nie znika. Nie ma jednego widocznego przełożonego, ale istnieje system określający tempo, kolejność i opłacalność pracy.

Taka sytuacja jest szczególnie trudna dla konfucjańskiej etyki ról. Jeśli przełożony jest osobą, można wymagać od niego ren, yi i odpowiedzialności za podwładnego. Gdy decyzję prezentuje system, pojawia się ryzyko moralnej depersonalizacji: „to nie ja, tak obliczył algorytm". Algorytm nie może jednak przejąć odpowiedzialności moralnej, gdyż nie posiada statusu uczestnika wspólnoty zobowiązań. Odpowiedzialność musi zatem zostać rozdzielona pomiędzy projektantów, kierownictwo, operatora platformy, podmiot wdrażający i organy nadzoru.

Konfucjańska etyka ról może być zatem niezwykle wartościowa jako krytyka „bezpodmiotowej odpowiedzialności". Rola społeczna nie znika dlatego, że jej zadanie zostało technicznie delegowane. Dyrektor nie przestaje odpowiadać za system rekrutacyjny tylko dlatego, że pierwszej selekcji dokonuje model. Lekarz nie zostaje zwolniony z obowiązku troski przez fakt, że rekomendację wygenerował system diagnostyczny. Państwo nie może uznać decyzji administracyjnej za normatywnie neutralną tylko dlatego, że wyliczył ją algorytm. Delegacja obliczenia nie jest delegacją odpowiedzialności moralnej.

Sandelowski argument o moralnych granicach rynku można rozszerzyć analogicznie: przekazanie wyboru AI nie zmienia natury dobra, o którym system decyduje. Jeśli algorytm dystrybuuje dostęp do kredytu, pracy, edukacji czy opieki, nadal rozstrzygamy o dobrach społecznych. Techniczna mediacja nie czyni ich mniej moralnymi. Wręcz przeciwnie – zwiększa potrzebę jawnego określenia ich telosu, ponieważ decyzja podejmowana na wielką skalę powtarza ukryte założenia miliony razy. Algorytmiczne rządzenie posiada bowiem cechę, której brakuje pojedynczemu błędowi urzędnika: skalowalność normatywnego błędu. Niesprawiedliwy człowiek może skrzywdzić ograniczoną liczbę osób; niesprawiedliwa reguła uderza w wielu, a niesprawiedliwy model robi to automatycznie, szybko i z pozorem statystycznego obiektywizmu.

Właśnie dlatego kwestia jakości danych, dyskryminacji i walidacji modeli stała się elementem regulacji generatywnej AI w Chinach. Przepisy wymagają m.in. środków przeciw dyskryminacji ze względu na pochodzenie etniczne, wyznanie, narodowość, region, płeć, wiek, zawód czy zdrowie oraz zwiększenia przejrzystości i wiarygodności treści. Jednak nawet formalnie „niedyskryminujący" model może utrwalać struktury społeczne, jeśli uczy się na danych będących zapisem dawnych nierówności.

Odpowiedzialność za skutek i ochrona wspólnej rzeczywistości

Tutaj powraca problem, który analizowaliśmy przy kwestii płci. System może nie posiadać zmiennej „kobieta", a mimo to wykorzystywać cechy silnie skorelowane z tą kategorią. Może nie „chcieć" dyskryminować – ponieważ nie posiada intencji – lecz jego wynik może systematycznie różnicować grupy. Moralność intencji, tak istotna dla etyki cnoty, okazuje się niewystarczająca wobec systemów statystycznych. Potrzebna jest moralność skutku i struktury.

Zderzają się tu w nowej postaci Mencjusz i Han Fei. Pierwszy pytał o dobrą dyspozycję człowieka; drugi o system odporny na jego interesowność. AI ujawnia trzeci problem: nie wystarczy ani dobra intencja projektanta, ani formalny standard, jeśli złożony system produkuje skutki, których żaden pojedynczy uczestnik nie przewidział. Pojawia się konieczność ciągłego audytu rezultatów, a nie tylko oceny pierwotnych założeń. To właśnie dlatego regulacje algorytmiczne w Chinach wymagają regularnego przeglądu mechanizmów, modeli, danych i efektów ich zastosowania. Filozoficznie jest to znaczące przesunięcie: odpowiedzialność nie kończy się w chwili zatwierdzenia reguły. System uczący się i działający w zmiennym środowisku wymaga trwałego nadzoru. Fa przestaje być jednorazowo zapisanym standardem, a zaczyna przyjmować charakter regulacji systemu adaptacyjnego.

Inny problem pojawia się wraz z generatywną AI. O ile system rekomendacyjny organizuje istniejące informacje, o tyle model generatywny może tworzyć nowe. W konsekwencji pytanie normatywne nie dotyczy już tylko tego, co użytkownik zobaczy, ale czy będzie potrafił rozpoznać ontologiczny status treści. Chińskie przepisy o oznaczaniu treści generowanych i syntetyzowanych przez AI, obowiązujące od 1 września 2025 roku, wymagają w określonych sytuacjach zarówno oznaczeń jawnych, dostępnych użytkownikowi, jak i zapisów w metadanych; regulują one także obowiązki platform rozpowszechniających takie materiały.

Można to interpretować jako próbę ochrony epistemicznego li – społecznej konwencji pozwalającej odróżnić rodzaje wypowiedzi. Społeczeństwo potrzebuje nie tylko informacji, ale reguł wskazujących, czym dana treść jest: reklamą, opinią, dokumentem, fikcją, fotografią czy materiałem syntetycznym. Kiedy AI zaciera te granice, problemem przestaje być wyłącznie prawda pojedynczego zdania. Zagrożona zostaje infrastruktura, dzięki której uczestnicy życia publicznego wiedzą, jak interpretować przekaz. Sandelowski ideał deliberacji zakłada wspólny świat faktów, wystarczająco stabilny, by można było spierać się o ich moralne znaczenie. Jeśli strony tracą możliwość rozróżnienia źródła, autorstwa czy statusu materiału, spór polityczny i moralny

przesuwa się z poziomu interpretacji na poziom samej rzeczywistości. Generatywna AI może zatem wzmacniać nie tylko dezinformację, lecz epistemiczną fragmentację wspólnoty.

Do tej fragmentacji przyczyniają się również systemy personalizacji. Jeśli każdy obywatel otrzymuje odmienny strumień informacji, zanika część wspólnego doświadczenia, które Sandel uważa za warunek republikańskiej wspólnoty. Problem skyboxification z poprzedniej części uzyskuje wersję cyfrową. Nie trzeba już mieszkać w innych dzielnicach i chodzić do różnych szkół, by żyć w odmiennych światach; wystarczy, że algorytmy uczą się naszych zainteresowań i utwierdzają nas w odrębnych środowiskach informacyjnych.

Z drugiej strony personalizacja może zwiększać dostęp do istotnych informacji, redukować przeciążenie poznawcze i pomagać w odnajdywaniu niszowych wspólnot. Nie istnieje więc prosty argument: „algorytm rekomendacyjny równa się fragmentacja". Kluczowe są parametry systemu: czy optymalizuje on różnorodność, czas zaangażowania, zgodność z profilem, jakość źródła, czy może kilka celów jednocześnie. Znów wracamy do telosu. W tym punkcie he er bu tong może zostać przepisane na język cyfrowy.

Harmonia bez jednolitości wymaga systemu informacyjnego, który pozwala ludziom posiadać różne poglądy, lecz nie zamyka ich całkowicie w odrębnych rzeczywistościach. Zbyt silna homogenizacja informacji prowadzi ku tong – jednolitości. Zbyt silna personalizacja może natomiast zniszczyć he – wspólną przestrzeń wzajemnego odniesienia. Dobry system potrzebuje zarazem różnicy i kontaktu między różnicami.

Predykcja danych nie może zastąpić prawa człowieka do zmiany i transformacji

W tym miejscu daoistyczny perspektywizm Zhuangziego przestaje być jedynie historyczną ciekawostką. Koncepcja qi wu lun służy tu jako przypomnienie o sytuacyjnym uwarunkowaniu naszych perspektyw i konieczności wyjścia poza binarny podział na „to" i „tamto". Algorytm personalizacyjny może jednak działać w przeciwnym kierunku: zamiast uświadamiać nam częściowość własnego spojrzenia, dostarcza treści, które jedynie je potwierdzają. Podczas gdy Zhuangzi zachęca do poznawczej decentracji, systemy rekomendacyjne wzmacniają centrum naszego cyfrowego profilu. Dlatego zadaniem AI nie powinno być wyłącznie coraz precyzyjniejsze „poznawanie użytkownika". Paradoksalnie, system naprawdę wspierający autonomię powinien czasami prezentować człowiekowi treści, których model nie przewiduje na podstawie jego

dotychczasowych wyborów. W przeciwnym razie przeszłość staje się cyfrowym przeznaczeniem, a predykcja zaczyna konserwować tożsamość.

Ma to kluczowe znaczenie dla Sandelowskiego „ja refleksyjnie usytuowanego". Jeśli jesteśmy w części ukształtowani przez historię, lecz zachowujemy zdolność do refleksji nad nią, system cyfrowy nie może traktować historii zachowań jako zamkniętej definicji osoby. Prawo do zmiany jest immanentną cechą podmiotowości; człowiek nie jest sumą poprzednich kliknięć.

Tutaj ujawnia się głęboka różnica między konfucjańską samokultywacją a cyfrową personalizacją. Samokultywacja zakłada możliwość stawania się innym człowiekiem, podczas gdy personalizacja oparta na ekstrapolacji danych zakłada przede wszystkim prawdopodobieństwo pozostania podobnym do siebie. Pierwsza jest teorią transformacji, druga – predykcji. Dopóki system rekomendacyjny jedynie wspiera aktualne preferencje, nie ma sprzeczności. Problem pojawia się w momencie, gdy zdolność przewidywania ogranicza ekspozycję na możliwości, dzięki którym te preferencje mogłyby ewoluować. Z tego względu wolność w epoce AI wymaga czegoś więcej niż tylko opcji „wyłącz personalizację". Republikańska teoria wolności jako niedominacji pytałaby, czy człowiek pozostaje zależny od władzy zdolnej arbitralnie kształtować jego otoczenie. Konfucjanizm zapytałby, czy środowisko wspiera właściwy rozwój relacji. Sandel zapytałby, jakiemu dobru społecznemu służy system. Daoizm zaś – czy człowiek wciąż potrafi wyjść poza mechanizm, który nieustannie interpretuje i przewiduje jego zachowanie.

W ten sposób wyłaniają się cztery odmienne, lecz komplementarne testy technologii. Test liberalny bada prawa jednostki, zgodę, prywatność i możliwość odmowy. Test republikański analizuje niedominację, kontrolę nad instytucjami oraz publiczną odpowiedzialność. Test konfucjański ocenia jakość relacji oraz charakter osób i organizacji wdrażających technologię. Test daoistyczny bada stopień instrumentalizacji, nadmierną kontrolę i utratę spontaniczności. Żaden z nich nie jest wystarczający w izolacji.

Chiński reżim regulacyjny AI odzwierciedla złożoność tych problemów. Obejmuje on ustawę o ochronie danych osobowych, przepisy o rekomendacjach algorytmicznych, regulacje dotyczące głębokiej syntezy, tymczasowe zasady generatywnej AI oraz – od 2025 roku – szczegółowe obowiązki oznaczania treści syntetycznych. Jednocześnie plan legislacyjny Rady Państwa na 2026 rok zapowiada dalsze prace nad kompleksową regulacją AI, obejmującą dane, moc obliczeniową, algorytmy, własność, cyberbezpieczeństwo, łańcuchy dostaw i kluczowe zastosowania. Oznacza to, że we wrześniu 2026 roku chiński model pozostaje przede wszystkim warstwowym systemem regulacji sektorowych i technologicznych, znajdując się jednocześnie w procesie dalszej konsolidacji legislacyjnej.

Odpowiedzialność człowieka w architekturze algorytmicznej

Ta warstwowość jest sama w sobie interesująca filozoficznie, gdyż AI nie stanowi jednego, spójnego przedmiotu. Rekomendacja treści, automatyczna rekrutacja, modele językowe, systemy diagnostyczne czy algorytmy sterujące pracą kuriera generują odmienne ryzyka. Próba uregulowania tych wszystkich obszarów jedną abstrakcyjną kategorią mogłaby być równie nieadekwatna, co próba opisania wszelkich relacji społecznych jednym rytuałem. Xunziańska wrażliwość na funkcjonalne różnicowanie ról znajduje tu nieoczekiwane echo: regulacja musi uwzględniać konkretną praktykę i jej cel. Jednocześnie nadmierna fragmentacja przepisów może utrudnić obywatelowi ustalenie, jakie prawa przysługują mu w starciu z konkretnym systemem. Dlatego przyszłość governance AI będzie zależeć nie tyle od liczby regulacji, co od ich zdolności do stworzenia czytelnej architektury odpowiedzialności.

Osobom dotkniętym błędem nie wystarczy wiedza o tym, że system składa się z dostawcy modelu, integratora, administratora danych, platformy i operatora. Musi istnieć konkretny podmiot, wobec którego można skutecznie wysunąć roszczenie. To właśnie jest granica moralnej metafory mówiącej, że „AI podejmuje decyzję". AI generuje wynik, lecz to społeczeństwo decyduje o wadze tego wyniku. Model może rekomendować odmowę, ale to instytucja rozstrzyga, czy ta rekomendacja stanie się wiążącą decyzją. System może oszacować ryzyko, jednak to człowiek lub prawo określają próg działania. Odpowiedzialność moralna zaczyna się dokładnie tam, gdzie organizacja postanawia przekształcić predykcję w konsekwencję.

W tym ujęciu AI nie zastępuje fa – staje się nowym przedmiotem fa i jednocześnie nowym instrumentem władzy. Nie zastępuje też li, ponieważ sposób korzystania z technologii wciąż zależy od kultury organizacyjnej, etosu zawodu i norm nieformalnych. Nie eliminuje ren, gdyż człowiek nadal może zostać potraktowany albo humanitarnie, albo instrumentalnie. I nie znosi yi, ponieważ realne sytuacje wciąż wymagają osądu, którego nie da się całkowicie zredukować do uniwersalnej funkcji celu. Największe niebezpieczeństwo pojawia się wtedy, gdy trzy porządki – li, fa i kod – zaczynają wzajemnie wzmacniać tę samą normę bez możliwości zewnętrznej korekty. Jeśli społeczeństwo moralnie piętnuje określone zachowanie, prawo je sankcjonuje, a system technologiczny utrudnia jego realizację lub widoczność, przestrzeń refleksyjnego dystansu jednostki dramatycznie się kurczy. Z drugiej strony, gdy te trzy porządki kontrolują się wzajemnie, mogą zwiększać poziom ochrony: kultura odpowiedzialności ogranicza oportunizm projektanta, prawo hamuje arbitralność organizacji, a technologia pomaga wykrywać nadużycia. AI nie powinna być zatem analizowana jako czwarta władza stojąca obok społeczeństwa, lecz jako wzmacniacz

sposobu, w jaki to społeczeństwo definiuje dobro, ryzyko, produktywność i normalność. System nie musi posiadać własnej ideologii, by zwielokrotniać tę wpisaną w dane, cel i instytucję wdrażającą.

Daoistyczne jixin stanowi tutaj ostrzeżenie przed podwójną redukcją. Najpierw świat zostaje przełożony na dane, a następnie człowiek zaczyna wierzyć, że to, czego nie udało się zdigitalizować, ma mniejsze znaczenie. Wskaźnik staje się celem, model rzeczywistością, a predykcja zostaje pomylona z wiedzą o człowieku. Wtedy technologia przestaje być narzędziem i staje się epistemologią. Konfucjanizm odpowiada na tę redukcję przypomnieniem, że osoba jest procesem moralnym, a nie rekordem danych. Sandel wskazuje, że dobra społeczne mają telos, którego nie można wywieść z funkcji optymalizacyjnej. Liberalizm podkreśla, że jednostka posiada roszczenia, których efektywność systemu nie może automatycznie unieważnić. Republikanizm zaś przypomina, że władza musi pozostać kontrolowalna, nawet jeśli operuje poprzez model zamiast urzędnika.

Mądrość wyznaczania granic optymalizacji AI

Daoizm stawia pytanie najprostsze, a zarazem najtrudniejsze: czy każda rzecz, którą potrafimy przewidzieć, zmierzyć i kontrolować, powinna zostać poddana tym procesom? To właśnie tutaj zhi zhi – wiedza o momencie zatrzymania – staje się jedną z kluczowych cnót cywilizacji sztucznej inteligencji. Rozwój AI wymaga nie tylko potężniejszych modeli, większych zbiorów danych czy mocy obliczeniowej, lecz przede wszystkim społecznej zdolności do uznania, że pewnych typów optymalizacji nie należy prowadzić, nawet jeśli są one technicznie możliwe. Nie każda decyzja powinna być automatyzowana, nie każda relacja profilowana, a nie każda niepewność jest błędem wymagającym usunięcia. Właśnie tu przebiega najgłębsza różnica między inteligentnym systemem a mądrym ładem społecznym. System potrafi coraz trafniej przewidzieć, jakie działanie doprowadzi do wskazanego wyniku. Mądry porządek musi natomiast wiedzieć, którego wyniku warto pragnąć, jakich środków nie wolno używać i jaka część ludzkiego życia powinna pozostać otwarta na nieprzewidywalność. AI może zwiększać pierwszą z tych zdolności; drugiej nie wolno jej milcząco przypisać.

Współczesne Chiny stają się w tym kontekście szczególnie istotnym laboratorium nowoczesności. Spotykają się tam silna tradycja moralnej formacji człowieka, legistyczne doświadczenie administracyjnego standardu, nowoczesne państwo regulacyjne i ogromna gospodarka platformowa z szybko rozwijaną AI oraz daoistycznym repertuarem krytyki nadmiernego sterowania. Nie tworzy to jednej, harmonijnej doktryny, lecz niezwykle gęsty obszar konfliktu między cnotą, kontrolą, efektywnością, prawami, relacyjnością a technologiczną możliwością.

Synteza nie może sprowadzać się ani do postulatu „więcej technologii”, ani do naiwnego „powrotu do tradycji”. Problem jest głębszy.

Cywilizacja cyfrowa potrzebuje pojęcia osoby, którego nie da się zredukować do użytkownika, pracownika, konsumenta czy profilu ryzyka; pojęcia prawa, które reguluje kod, nie stając się samym kodem; kultury moralnej, która wychowuje bez uniformizacji, oraz teorii wspólnoty zachowującej możliwość sporu. Dopiero połączenie tych elementów pozwala przejść od sztucznej inteligencji jako technologii przewidywania do społecznej mądrości zdolnej wyznaczać granice tego procesu. Wniosek ten prowadzi bezpośrednio do problemu, który domyka konstrukcję artykułu: pluralizmu bez pustki i wspólnoty bez uwięzienia jednostki. Po analizie Konfucjusza, Xunziego, tradycji fa, daoizmu, Sandela, harmonii, merytokracji, edukacji, płci, genetyki, rynku, prawa i AI pozostaje pytanie syntetyczne: jak zbudować społeczeństwo, które posiada moralną treść, lecz nie narzuca jednej obowiązkowej definicji dobrego życia; które uznaje człowieka za istotę relacyjną, ale pozwala mu odrzucić złą relację; które potrzebuje norm, ale nie myli harmonii z ciszą; i które wykorzystuje technologię, nie pozwalając jej określać, czym jest człowiek.

> Czy potrafimy jeszcze odróżnić świat od sposobu jego pomiaru, czy może staliśmy się już tylko sumą naszych kliknięć i predykcji? Prawdziwa mądrość cywilizacji AI nie będzie mierzona mocą obliczeniową modeli, lecz odwagą do pozostawienia części ludzkiego życia w sferze nieprzewidywalności. Ostatecznie to właśnie tam, gdzie kończy się optymalizacja a zaczyna wolność, definiuje się nasza podmiotowość.

Inspiracje

[1] "Encountering China" Michael J Sandel

Michael Joseph Sandel (ur. 5 marca 1953) jest amerykańskim filozofem politycznym i profesorem teorii rządu im. Anne T. i Roberta M. Bassów na Uniwersytecie Harvarda. Jest powszechnie uznawany za komunitariańską krytykę liberalnej teorii politycznej, szczególnie w odpowiedzi na „Teorię sprawiedliwości" Johna Rawlsa, argumentując, że jednostki są fundamentalnie osadzone w kontekstach społecznych i wspólnotowych. Sandel zyskał międzynarodową sławę dzięki swojej przystępnej filozofii publicznej, której przykładem jest jego legendarny kurs „Sprawiedliwość" na Harvardzie i powiązane z nim wykłady transmitowane na całym świecie. Jego badania naukowe dotyczą moralnych granic rynków, etyki biotechnologii oraz fragmentacji społecznej spowodowanej ideałami merytokratycznymi. Sandel, członek Amerykańskiej Akademii Sztuk i Nauk, jest autorem przełomowych prac na temat cnót obywatelskich, sprawiedliwości i dobra wspólnego.

Michael Joseph Sandel (born March 5, 1953) is an American political philosopher and the Anne T. and Robert M. Bass Professor of Government Theory at Harvard University. He is widely recognized for his communitarian critique of liberal political theory, particularly in response to John Rawls's A Theory of Justice, arguing that individuals are fundamentally embedded within social and communal contexts. Sandel gained international renown for his accessible public philosophy, exemplified by his legendary Harvard course 'Justice' and related lectures broadcast globally. His scholarship investigates the moral limits of markets, the ethics of biotechnology, and the social fragmentation caused by meritocratic ideals. A member of the American Academy of Arts and Sciences, Sandel has authored landmark works on civic virtue, justice, and the common good.

Harvard University

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:



[1] "Remix"

Lawrence Lessig

A&C Black | ISBN: 9781408113936



[2] "The Analects of Confucius"

Konfucjusz

ISBN: 9781636001227



[3] "Narrative Capitalism, Algorithmic Governance and Psychic Integrity"

Northon Salomão de Oliveira

2026 | Northon Salomão de Oliveira

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "The Subject and Power"

Michel Foucault

1982 | Critical Inquiry | DOI: 10.1086/448181

Źródło

[2] "Power/Knowledge: Selected Interviews and Other Writings 1972-1977"

Michel Foucault, Colin Gordon

1980

Źródło

[3] "The Order of Things: An Archaeology of the Human Sciences."

F. C. T. Moore, Michel Foucault

1971 | Man | DOI: 10.2307/2799252

Źródło

[4] "Code and Other Laws of Cyberspace"

Lawrence Lessig

1999 | Medical Entomology and Zoology

Źródło

[5] "Free culture: how big media uses technology and the law to lock down culture and control creativity"

Lawrence Lessig

2004 | Choice Reviews Online | DOI: 10.5860/choice.42-1641

Źródło

[6] **"Code: Version 2.0"**
Lawrence Lessig
2006
Źródło

[7] **"The First Amendment Does Not Protect Replicants"**
L. Lessig
2021 | Social Science Research Network | DOI: 10.2139/ssrn.3922565
Źródło

[8] **"Space, Knowledge, and Power [1982]"**
Michel Foucault
2023 | The Design Culture Reader | DOI: 10.4324/9781003416692-13
Źródło

Artykuły naukowe autora:

[9] **"Liberalism and the Limits of Justice"**
Michael J. Sandel
1998 | Cambridge University Press eBooks | DOI: 10.1017/cbo9780511810152
Źródło

[10] **"Democracy's Discontent: America in Search of a Public Philosophy."**
Paul K. Conkin, Michael J. Sandel
1997 | The American Historical Review | DOI: 10.2307/2170809
Źródło

[11] **"What Money Can't Buy: The Moral Limits of Markets"**
Michael J. Sandel
2012
Źródło

[12] **"The Procedural Republic and the Unencumbered Self"**
Michael J. Sandel
1984 | Political Theory | DOI: 10.1177/0090591784012001005
Źródło

[13] "The Case against Perfection"

Michael J. Sandel

2007 | Harvard University Press eBooks | DOI: 10.4159/9780674043060

Źródło

[14] "Liberalism and its critics"

Michael J. Sandel

1984 | New York University Press eBooks

Źródło

[15] "Justice: What's the Right Thing to Do?"

Michael J. Sandel

2009 | bepress Legal Repository

Źródło

[16] "Public Philosophy: Essays on Morality in Politics"

Michael J. Sandel

2005

Źródło

[17] "Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy"

Michael J. Sandel

2013 | The Journal of Economic Perspectives | DOI: 10.1257/jep.27.4.121

Źródło

[18] "Populism, liberalism, and democracy"

Michael J. Sandel

2018 | Philosophy & Social Criticism | DOI: 10.1177/0191453718757888

Źródło

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 7843c7b7-f0b4-4d52-b30f-be40166de250