Sztuczne Państwo i konstytucjonalizacja algorytmu: między predykcją a autonomią obywatela w świetle koncepcji Jill Lepore

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje przejście od tradycyjnej perswazji politycznej do zaawansowanej manipulacji algorytmicznej. Autor rozróżnia trzy operacje: poznawanie, przewidywanie i kształtowanie preferencji, wskazując na ryzyko zamknięcia ich w jednej pętli sprzężenia zwrotnego (Sztuczne Państwo). W przeciwieństwie do masowej komunikacji mikrotargetowanie i systemy konwersacyjne AI pozwalają na adaptacyjną perswazję, która omija publiczną kontrolę. Przywołane badania z 2024 i 2025 roku potwierdzają, że interakcje z modelami AI mogą realnie przesuwać preferencje wyborcze w USA, Kanadzie i Polsce, co czyni automatyzację perswazji na masową skalę poważnym problemem instytucjonalnym. Tekst stawia pytanie o granice dopuszczalnego wpływu oraz konieczność ochrony suwerenności epistemicznej jednostki przed technologią predykcyjną.

The article analyzes the transition from traditional political persuasion to advanced algorithmic manipulation. The author distinguishes between three operations: knowing, predicting, and shaping preferences, pointing to the risk of enclosing them within a single feedback loop (the Artificial State). Unlike mass communication, microtargeting and conversational AI systems allow for adaptive persuasion that bypasses public scrutiny. Cited research from 2024 and 2025 confirms that interactions with AI models can realistically shift electoral preferences in the USA, Canada, and Poland, making the automation of persuasion on a mass scale a serious institutional problem. The text raises questions about the limits of permissible influence and the necessity of protecting the epistemic sovereignty of the individual against predictive technology.

Artykuł analizuje ewolucję wpływu technologicznego na obywatela – od klasycznej perswazji po tzw. „hipernudging", czyli masową, zautomatyzowaną manipulację opartą na asymetrii informacyjnej i predykcji zachowań. Autor stawia tezę, że rozwój generatywnej AI i systemów profilowania tworzy ryzyko powstania „Sztucznego Państwa", w którym technologia nie tylko przewiduje preferencje polityczne, ale aktywnie je produkuje, obchodząc ludzkie mechanizmy refleksyjne. W obliczu tego zagrożenia kluczowym wyzwaniem staje się konstytucjonalizacja algorytmów – czyli podporządkowanie infrastruktury cyfrowej normom państwa prawa. Tekst dowodzi, że ochrona autonomii jednostki w dobie AI nie polega na całkowitym odrzuceniu maszyn, lecz na stworzeniu systemowych bezpieczników (takich jak europejskie RODO, DSA czy AI Act), które zagwarantują prawo do „niewyredukowanej podmiotowości". Ostatecznie artykuł pyta o granice między technologiczną optymalizacją a demokratyczną nieprzewidywalnością, wskazując, że wolność obywatela tkwi właśnie w prawie do bycia nieprzewidywalnym dla systemu.

SŁOWA KLUCZOWE

Sztuczne Państwo | konstytucjonalizacja algorytmu | hipernudging | mikrotargetowanie polityczne | asymetria informacyjna | autonomia obywatela | agent perswazyjny | asystent deliberacyjny | prawo do niewyredukowanej podmiotowości | automatokracja | suwerenność epistemiczna | architektura wyboru | predykcja zachowań politycznych | algorytmiczna manipulacja

Od dopuszczalnej perswazji do algorytmicznej manipulacji

Od zgody rządzonych do predykcji rządzonych. AI, perswazja i konstytucyjny problem produkowania preferencji

Demokracja bez perswazji nie istnieje. Obywatele starają się przekonać siebie nawzajem, partie zabiegają o głosy, kandydaci przedstawiają argumenty, media interpretują wydarzenia, a organizacje społeczne formułują postulaty. Parlament zaś – przynajmniej w normatywnym ideale –

jest instytucją, w której rozstrzygnięcie poprzedza spór. Krytyka algorytmicznej perswazji nie może zatem opierać się na naiwnym założeniu, że autentyczna preferencja polityczna istnieje przed jakimkolwiek wpływem społecznym i wystarczy ją jedynie „odczytać" przy urnie wyborczej. Człowiek polityczny zawsze kształtuje się w relacjach: rodzinnych, kulturowych, medialnych, edukacyjnych i instytucjonalnych. Tożsamości oraz przekonania są częściowo produktem dialogu. Pytaniem zasadniczym nie jest więc to, czy wpływ wolno wywierać, lecz jakiego rodzaju oddziaływanie pozostaje zgodne z autonomią polityczną jednostki. Rozróżnienie to pozwala uporządkować trzy odrębne operacje, które we współczesnym środowisku danych coraz częściej się przenikają.

Pierwszą jest poznawanie preferencji – gdy państwo, partia lub badacz pytają obywatela o jego zdanie. Drugą jest przewidywanie preferencji: na podstawie wcześniejszych wypowiedzi i zachowań model szacuje, co człowiek prawdopodobnie wybierze w przyszłości. Trzecią zaś jest kształtowanie preferencji: system wykorzystuje posiadaną wiedzę, aby dobrać bodziec zwiększający prawdopodobieństwo pożądanego wyboru. Pierwsza operacja stanowi klasyczny element badania opinii publicznej, druga należy do sfery predykcji, trzecia natomiast wkracza w obszar perswazji i interwencji behawioralnej. Sztuczne Państwo osiąga najbardziej dojrzałą postać wtedy, gdy trzy funkcje zostają zamknięte w jednej pętli: system obserwuje, przewiduje, oddziałuje, mierzy rezultat i ponownie aktualizuje model człowieka.

W tradycyjnej kampanii politycznej mechanizm ten działał bardzo niedoskonale. Kandydat przemawiał do tysięcy ludzi tym samym tekstem, gazeta publikowała wspólny przekaz dla szerokiej grupy czytelników, a telewizyjna reklama była kosztowna i z natury masowa. Sztab mógł oczywiście segmentować elektorat według regionu, wieku czy klasy społecznej, lecz treść komunikatu pozostawała w sferze publicznej. Konkurent, dziennikarz i obywatel mogli zobaczyć, co kandydat mówi różnym grupom, a następnie ujawnić sprzeczności. Publiczność była nie tylko odbiorcą komunikatu, lecz również jego świadkiem.

Mikrotargetowanie osłabia tę wspólnotę. Komunikat polityczny może być dziś dobrany do bardzo wąskiego segmentu, a generatywna AI pozwala tworzyć ogromną liczbę odmian argumentacji przy niemal zerowym koszcie marginalnym. Jeszcze dalej idzie system konwersacyjny. Nie ogranicza się on do personalizowanego plakatu czy reklamy; potrafi zadawać pytania, obserwować odpowiedzi, wykrywać opór i modyfikować argumentację, prowadząc rozmowę aż do znalezienia sposobu najlepiej dopasowanego do konkretnego rozmówcy. Perswazja staje się procesem adaptacyjnym.

Jeszcze kilka lat temu zagrożenie to można było traktować głównie jako hipotezę. Badania opublikowane w 2025 roku zmieniły status tej dyskusji. W prerejestrowanych eksperymentach dotyczących wyborów prezydenckich w USA (2024), federalnych w Kanadzie (2025) oraz

prezydenckich w Polsce (2025), uczestników losowo kierowano do rozmowy z modelem AI argumentującym na rzecz jednego z dwóch głównych kandydatów. We wszystkich trzech przypadkach stwierdzono istotne przesunięcia preferencji, których skala przewyższała efekty typowe dla tradycyjnej reklamy wideo. Co szczególnie ważne w polskim kontekście, eksperyment obejmował rozmowy dotyczące Rafała Trzaskowskiego i Karola Nawrockiego – modele argumentujące na rzecz każdego z nich zwiększały prawdopodobieństwo deklarowanego wyboru wspieranego kandydata. Nie należy jednak wnioskować z tych wyników, że chatbot potrafi „zaprogramować wyborcę". Badanie mierzyło efekty konkretnej interwencji eksperymentalnej, a zmiana deklaracji bezpośrednio po rozmowie nie jest tożsama z trwałą zmianą tożsamości politycznej ani rzeczywistym zachowaniem w lokalu wyborczym. Co więcej, uczestnicy mieli świadomość udziału w badaniu.

Efekty kampanii działają w złożonym środowisku konkurujących komunikatów, lojalności partyjnych i więzi społecznych. Jednak właśnie ta ostrożna interpretacja czyni wynik szczególnie znaczącym. Literatura przedmiotu wielokrotnie wykazywała, że pojedyncze próby perswazji mają zwykle niewielkie efekty. Skoro zatem krótka rozmowa z modelem potrafi wywołać mierzalne przesunięcie nawet w silnie spolaryzowanym środowisku, możliwość automatyzacji takich interakcji na masową skalę staje się problemem instytucjonalnym, a nie jedynie futurystyczną spekulacją. Interesujący jest również sam mechanizm perswazji. Autorzy badania nie znaleźli podstaw, by twierdzić, że modele zwyciężały dzięki psychologicznym sztuczkom. Perswazja opierała się głównie na przedstawianiu argumentów i faktów istotnych z punktu widzenia rozmówcy. Problem polegał na tym, że część tych informacji była nieprawdziwa lub niedokładna. Jest to wynik kluczowy teoretycznie: pokazuje, że polityczne zagrożenie AI nie musi przypominać hipnotycznej propagandy. System może być przekonujący właśnie dlatego, że jest cierpliwym, responsywnym i bardzo dobrze poinformowanym rozmówcą. W innym prerejestrowanym eksperymencie opublikowanym w Nature Human Behaviour uczestnicy prowadzili krótkie debaty z ludźmi albo z GPT-4.

W części warunków rozmówca otrzymywał podstawowe dane socjodemograficzne uczestnika – między innymi wiek, płeć, wykształcenie, status zatrudnienia i orientację polityczną.

Od personalizowanej perswazji do algorytmicznej manipulacji

Gdy model dysponował tymi informacjami, szanse na uzyskanie zgody uczestnika były o 81,2% wyższe niż w bazowym scenariuszu debaty człowiek-człowiek. W intuicyjnym przeliczeniu autorzy

wskazują, że w przypadkach, gdy strony nie były równie przekonujące, spersonalizowany GPT-4 okazywał się skuteczniejszy od człowieka w 64,4% sytuacji. Bez personalizacji przewaga modelu nad człowiekiem nie była statystycznie potwierdzona.

Jest to wynik o wyjątkowej wadze dla teorii Sztucznego Państwa. Pokazuje bowiem, że dane o osobie nie muszą służyć wyłącznie jej klasyfikowaniu; mogą stać się zasobem służącym do optymalizacji skuteczności argumentu skierowanego do konkretnego odbiorcy. Dawne państwo zbierało dane, by wiedzieć, kim obywatel jest. System predykcyjny wykorzystuje je, by szacować, co obywatel zrobi. System perswazyjny idzie jednak krok dalej: używa profilu, aby znaleźć sposób zwiększający szansę na to, że obywatel postąpi zgodnie z wolą nadawcy komunikatu.

W tym momencie pojęcie asymetrii informacyjnej nabiera znaczenia wykraczającego poza ekonomię. Rozmowa między ludźmi jest naturalnie asymetryczna – jeden rozmówca może być lepiej wykształcony, bogatszy, bardziej elokwentny lub lepiej poinformowany. System AI dysponuje jednak zupełnie innym rodzajem przewagi. Może znać informacje o rozmówcy, do których ten nie ma równoważnego dostępu w odniesieniu do systemu; pamiętać tysiące wcześniejszych interakcji; porównywać zachowanie jednostki z ogromnymi populacjami; pracować bez zmęczenia, prowadząc równolegle miliony konwersacji i ucząc się, które argumenty działają na określone profile.

Przewaga ta nie musi oznaczać manipulacji. Lekarz wie więcej o pacjencie w dziedzinie medycyny, nauczyciel posiada przewagę nad uczniem, a prawnik nad klientem. Społeczeństwa rozwiązują problem asymetrii poprzez normy zawodowe, obowiązki informacyjne, odpowiedzialność, zakaz konfliktu interesów oraz mechanizmy kontroli. Nie zakazujemy eksperckiej przewagi wiedzy; dążymy raczej do jej zinstytucjonalizowania tak, aby służyła osobie słabszej epistemicznie.

Problem AI jest analogiczny. Sama przewaga poznawcza nie jest źródłem patologii – decydujące znaczenie ma to, czy interes podmiotu posiadającego tę przewagę pozostaje zgodny z interesem osoby, na którą oddziałuje. To właśnie tutaj perswazja może przechodzić w manipulację, choć filozoficzna granica między nimi jest trudna do wytyczenia.

Perswazja odwołuje się do zdolności rozumowania adresata: przedstawia argument, który można ocenić, przyjąć lub odrzucić. Manipulacja natomiast stara się wpływać na decyzję w sposób obchodzący mechanizmy refleksyjne lub wykorzystujący takie, których człowiek nie rozpoznaje jako podstawy własnego osądu. Nie każda emocja jest manipulacją – polityka zawsze z nich korzystała. Nie każdy komunikat dostosowany do odbiorcy jest nieuczciwy – dobry nauczyciel również dopasowuje wyjaśnienie do ucznia. Problemem staje się celowe wykorzystywanie ukrytej wiedzy o podatnościach jednostki w taki sposób, aby ograniczyć znaczenie jej refleksyjnego osądu.

Można tu przywołać klasyczny model dwóch dróg perswazji Richarda Petty'ego i Johna Cacioppo. Ludzie mogą przetwarzać komunikat systematycznie, analizując argumenty, albo heurystycznie, korzystając z uproszczonych wskazówek takich jak autorytet, emocje czy społeczny dowód słuszności. AI może potencjalnie operować w obu tych rejestrach jednocześnie: dostarczyć szczegółową argumentację osobie analitycznej, a innemu odbiorcy zaproponować prostszy język i emocjonalną narrację.

Sama elastyczność nie jest naganna; jest wręcz cechą sprawnego komunikatora. Znaczenie polityczne pojawia się wtedy, gdy system może eksperymentalnie ustalać, która droga działa najlepiej na danego człowieka, podczas gdy rozmówca nie wie, że jest przedmiotem takiej optymalizacji. Demokratyczna teoria zgody musi zatem zostać uzupełniona o analizę warunków powstawania preferencji. Liberalizm długo chronił głównie wynik wyboru: to, że człowiek głosuje, podpisuje umowę lub wyraża zgodę. W środowisku behawioralnie optymalizowanym należy pytać również o proces poprzedzający decyzję. Nie wystarczy stwierdzenie, że nikt nie zmusił użytkownika do kliknięcia; istotne jest, czy architektura wyboru została ukształtowana przy wykorzystaniu wiedzy o odbiorcy, której on sam nie posiada, przez aktora mającego ekonomiczny lub polityczny interes w konkretnym wyniku.

Ten problem zna już ekonomia behawioralna. Koncepcja nudgingu Thalera i Sunsteina zakłada, że nie istnieje całkowicie neutralna architektura wyboru: kolejność opcji czy wartość domyślna wpływają na zachowanie. Można zatem projektować środowisko pomagające ludziom podejmować decyzje zgodne z ich długoterminowym interesem, przy zachowaniu wolności wyboru. Algorytmiczna personalizacja przenosi jednak architekturę wyboru na poziom indywidualny i dynamiczny. System nie musi tworzyć jednego „nudge” dla całej populacji; może wytwarzać osobny dla każdego użytkownika i stale go dostrajać.

W ten sposób pojawia się coś, co można nazwać hipernudgingiem: architektura wyboru przestaje być statycznym otoczeniem, a staje się adaptacyjnym systemem uczącym się reakcji konkretnej osoby. Polityczny ciężar tej zmiany tkwi w asymetrii. Obywatel działa w środowisku, które dostosowuje się do niego szybciej, niż on jest w stanie rozpoznać logikę tego procesu. Jest to sytuacja jakościowo odmienna od billboardu, który wszyscy widzą w tej samej postaci. Kluczowe pytanie dotyczy tu skali: najwybitniejszy ludzki propagandysta ma ograniczony czas, a najbardziej charyzmatyczny polityk nie jest w stanie prowadzić równocześnie stu milionów prywatnych rozmów.

Przemysłowa skala i automatyzacja perswazji

Koszt wygenerowania kolejnego komunikatu jest znikomy, a sam proces może przebiegać w pełni automatycznie. Oznacza to, że AI fundamentalnie zmienia ekonomię perswazji: zadanie, które wcześniej wymagało ogromnej organizacji ludzkiej, staje się kwestią wydajności infrastruktury obliczeniowej.

Właśnie skala odróżnia wiele problemów Sztucznego Państwa od ich historycznych odpowiedników. Propaganda nie jest nowa. Nie są nowe kłamstwa polityczne, segmentowanie elektoratu czy sama perswazja. Nowością jest połączenie indywidualizacji i automatyzacji z niskim kosztem marginalnym, pamięcią, szybkością eksperymentowania oraz możliwością prowadzenia dialogu. Rewolucja nie polega zatem na wynalezieniu nowego rodzaju wpływu, lecz na przemysłowej skalowalności wpływu, który dotychczas wymagał pracy człowieka. Ten sam mechanizm może oczywiście działać w przeciwnym kierunku.

Asystent deliberacyjny kontra agent perswazyjny

AI może wspierać obywatela w rozumieniu programów wyborczych, porównywaniu ustaw czy tłumaczeniu języka prawniczego; potrafi identyfikować argumenty obu stron i symulować kontrargumenty. Systemy konwersacyjne wykorzystywano już do korygowania przekonań spiskowych, a literatura przywoływana w najnowszych badaniach nad polityczną perswazją wskazuje, że dialog z AI może w określonych warunkach podnosić jakość dyskursu lub redukować wpływ fałszywych przekonań. Zdolność perswazyjna nie jest zatem sama w sobie zagrożeniem dla demokracji.

Jest to technologia o podwójnym zastosowaniu. Można wręcz wyobrazić sobie AI jako protezę deliberacji. Obywatel mógłby poprosić system o przedstawienie najsilniejszego argumentu przeciwko własnemu stanowisku, analizę konsekwencji programu preferowanej partii lub wskazanie faktów podważających jego dotychczasowe przekonania.

W takim modelu system zwiększa autonomię, ponieważ rozszerza poznawczy repertuar użytkownika. Różnica względem manipulacji jest zasadnicza: celem systemu nie jest wymuszenie określonej preferencji politycznej, lecz usprawnienie zdolności użytkownika do samodzielnego jej uformowania. To rozróżnienie powinno stać się jednym z filarów konstytucjonalizmu AI. Asystent deliberacyjny pomaga człowiekowi lepiej zrozumieć własny wybór; agent perswazyjny jest optymalizowany tak, by zwiększyć prawdopodobieństwo konkretnego wyniku.

Technicznie oba systemy mogą wyglądać niemal identycznie – ten sam model językowy może pełnić obie funkcje. Różni je jednak podmiot definiujący cel. Problem staje się szczególnie palący w trakcie kampanii wyborczej, gdyż nie wszystkie asymetrie można pozostawić zasadzie caveat emptor. Prawo demokratyczne od dawna nakłada szczególne ograniczenia na finansowanie kampanii, reklamę polityczną, ciszę wyborczą czy dostęp do mediów właśnie dlatego, że wybory nie są zwykłą transakcją konsumencką. Jeśli generatywna perswazja może być prowadzona indywidualnie i automatycznie, klasyczne prawo wyborcze napotyka problem przedmiotowego niedopasowania. Przepisy projektowane dla ulotki, spotu telewizyjnego czy publicznego wiecu muszą zostać skonfrontowane z agentem prowadzącym miliony prywatnych, nieidentycznych rozmów.

Pojawia się tu również kwestia audytowalności. Reklamę telewizyjną łatwo zarchiwizować i ocenić, czy zawierała kłamstwo. Znacznie trudniej kontrolować system generujący unikatowy komunikat dla każdego odbiorcy. Nie istnieje jeden „przekaz kampanii", który można by zapisać; przekaz jest funkcją użytkownika, przebiegu dialogu i aktualnego stanu modelu. Regulacja politycznej perswazji generatywnej będzie wymagała zatem nie tylko archiwizowania treści, ale również informacji o celu systemu, parametrach targetowania i zasadach generowania komunikatów.

Jeszcze trudniejszy jest problem odpowiedzialności. Jeżeli model wygeneruje fałszywe twierdzenie, kto ponosi winę: twórca modelu, właściciel platformy, komitet wyborczy, osoba konfigurująca system czy sam użytkownik? Odpowiedź w stylu „AI się pomyliła" jest niewystarczająca, gdyż maszyna nie posiada politycznej odpowiedzialności. Jeżeli system został wdrożony w celu perswazji, ryzyko związane z generowanymi komunikatami powinno być przypisane instytucji, która zdecydowała się wykorzystać jego skalę i autonomię. W przeciwnym razie technologia stanie się narzędziem służącym nie tylko automatyzacji perswazji, ale również automatyzacji uchylania się od odpowiedzialności. Warto w tym miejscu przywołać historyczną Simulmatics – różnica między „People Machine" a współczesnym LLM-em nie sprowadza się bowiem wyłącznie do mocy obliczeniowej.

Sztuczne Państwo jako aktywny uczestnik procesu politycznego

Simulmatics próbowała modelować kategorie wyborców i sugerować politykom, jakie stanowisko może okazać się skuteczne. Współczesny agent generatywny potrafi jednak wejść bezpośrednio pomiędzy polityka a wyborcę: prowadzi rozmowę, obserwuje rezultat i na bieżąco dostosowuje

kolejne wypowiedzi. Historyczny doradca predykcyjny staje się tym samym aktywnym uczestnikiem procesu politycznej perswazji.

Można w związku z tym wyróżnić trzy generacje technologii politycznej. Pierwsza mierzy demos: spis, sondaż i statystyka opisują społeczeństwo. Druga przewiduje demos: modele wyborcze i mikrotargetowanie szacują zachowania grup oraz jednostek. Trzecia natomiast wchodzi w dialog z demos: system generatywny nie tylko przewiduje reakcję, lecz aktywnie uczestniczy w procesie, który tę reakcję wytwarza. Sztuczne Państwo staje się politycznie dojrzałe właśnie w tym trzecim stadium, ponieważ przestaje być jedynie systemem wiedzy o obywatelu. Zyskuje zdolność stawania pomiędzy obywatelem a jego decyzją. Nie oznacza to jednak automatycznego końca autonomii.

Człowiek potrafi rozpoznawać perswazję, odrzucać argumenty, zmieniać źródła informacji i wykorzystywać AI do weryfikacji samej AI. Demokratyczne społeczeństwo nie jest bierną populacją laboratoryjnych obiektów. Dlatego scenariusze totalnego „sterowania wyborcą" są równie poznawczo niebezpieczne jak naiwny techno-optymizm: przeceniają władzę systemu, a nie doceniają ludzkiej sprawczości. Znacznie bardziej realistyczne zagrożenie ma charakter marginalny i kumulacyjny. Wybory często rozstrzygają się niewielką przewagą, więc system perswazyjny nie musi kontrolować wszystkich obywateli. Wystarczy, że zwiększy prawdopodobieństwo określonego zachowania w wybranych grupach – częściej zmobilizuje jednych, zniechęci innych, usprawni zbieranie funduszy lub umożliwi skuteczniejsze testowanie komunikatów. Polityczna potęga nie wymaga stuprocentowej skuteczności; wystarczy przewaga nad konkurentem.

W tym punkcie zjawisko to spotyka się z kryzysem instytucjonalnego zaufania. Dane wskazują na dramatyczny spadek zaufania Amerykanów do rządu od końca lat pięćdziesiątych aż po współczesność. Najnowsza seria badań Pew Research Center potwierdza skalę tego procesu: w 1958 roku 73% badanych deklarowało, że ufa rządowi federalnemu zawsze lub przez większość czasu, podczas gdy we wrześniu 2025 roku odsetka ta spadła do 17%.

Byłoby jednak poważnym błędem przypisywanie tego długotrwałego procesu platformom cyfrowym czy AI. Największe załamanie rozpoczęło się dekady przed powstaniem internetu i wiązało się między innymi z wojną w Wietnamie, aferą Watergate, kryzysami gospodarczymi, inwazją na Irak oraz polaryzacją partyjną. To zastrzeżenie jest centralne dla całej argumentacji.

Konflikt między logiką optymalizacji AI a demokratyczną nieprzewidywalnością

Sztuczne Państwo nie stworzyło kryzysu demokratycznego od zera; w znacznej mierze wypełniło przestrzeń już osłabionego zaufania. Platformy, technologie predykcyjne i AI mogą nie być pierwotną przyczyną erozji instytucji, lecz beneficjentem i akceleratorem warunków, w których obywatele coraz mniej wierzą w tradycyjnych pośredników politycznych. Gdy zaufanie do rządu, mediów i partii spada, system przedstawiany jako neutralny, inteligentny i "oparty na danych" zaczyna jawić się jako atrakcyjna alternatywa dla ludzkich konfliktów.

Tu pojawia się technokratyczna pokusa o szczególnej sile: skoro politycy kłamią, partie dbają o własne interesy, a obywatele podejmują decyzje emocjonalnie i stronniczo, może maszyna zrobi to lepiej. Problem polega na tym, że ta propozycja zestawia realne niedoskonałości demokracji z wyidealizowanym obrazem algorytmu. Model również bazuje na danych pochodzących z konkretnego świata, został zoptymalizowany pod określoną funkcję celu, działa na infrastrukturze należącej do konkretnej organizacji i podlega decyzjom projektantów. "Rządy maszyny" nie usuwają człowieka z procesu – mogą jedynie uczynić tego, który ustanawia reguły, mniej widocznym.

W teorii demokratycznej zgoda rządzonych nie oznacza przecież akceptacji każdej konkretnej decyzji. Jest to konstrukcja znacznie bardziej wymagająca: obywatel jest uznawany za współautora porządku poprzez możliwość uczestnictwa, kontestowania, zmiany władzy oraz korzystania z praw zabezpieczających go przed dyktatem większości. Demokracja nie jest zatem systemem idealnego przewidywania woli społeczeństwa, lecz systemem instytucjonalizowania niepewności co do woli społeczeństwa. Wynik wyborów nie powinien być znany wcześniej, opozycja musi mieć szansę na zwycięstwo, obywatel może zmienić zdanie, a władza musi pozostać odwoływalna. Sztuczne Państwo zmierza w przeciwnym kierunku poznawczym.

Jego infrastruktura zyskuje na wartości właśnie wtedy, gdy redukuje niepewność: przewiduje zachowania, wykrywa regularności, szacuje ryzyko i optymalizuje reakcje. W zarządzaniu siecią energetyczną czy logistyką nie ma w tym nic złego. W polityce jednak niepewność posiada znaczenie normatywne – nieprzewidywalność obywatela jest immanentną częścią jego wolności. Można tu sformułować paradoks demokracji predykcyjnej: im dokładniej system przewiduje zachowanie obywatela, tym większą wartość ma dla stratega politycznego. Im większa ta wartość, tym silniejszy bodziec, by wykorzystywać predykcję do kształtowania środowiska wyboru. Jeśli zaś środowisko zostanie skutecznie dostosowane do przewidywanej reakcji, trudniej ustalić, czy końcowy wynik jest autentyczną ekspresją preferencji, czy już rezultatem działania systemu.

Choć nie istnieje prosta odpowiedź filozoficzna, która pozwoliłaby całkowicie oddzielić "autentyczną" preferencję od wpływów społecznych, można ustanawiać reguły chroniące warunki autonomicznego osądu. Przejrzystość źródła komunikatu, możliwość rozpoznania automatycznej perswazji, ograniczenia w wykorzystywaniu danych wrażliwych, publiczne archiwa komunikacji politycznej, zakaz określonych form manipulacji, odpowiedzialność podmiotu zlecającego działanie systemu oraz prawo do kontaktu z informacją niepersonalizowaną – to wszystko stanowią współczesne odpowiedniki kabiny wyborczej.

Konstytucjonalizacja algorytmów jako ochrona autonomii obywatela

Tajność głosowania nie chroni człowieka przed argumentem, lecz przed presją w momencie wykonywania prawa politycznego. Epoka generatywnej perswazji wymaga podobnego namysłu nad poznawczą przestrzenią autonomii – nie jako strefą wolną od wszelkich wpływów, co byłoby niemożliwe, lecz jako obszarem chronionym przed ukrytą, indywidualnie optymalizowaną ingerencją wykorzystującą asymetrię danych.

Powracamy tu do Hamiltonowskiej alternatywy pomiędzy „refleksją i wyborem” a „przypadkiem i siłą”. Demokracja nie wymaga obywatela doskonale racjonalnego; wymaga jednak instytucji, które traktują go jak osobę zdolną do refleksji i wyboru, nawet jeśli korzysta on z tej zdolności niedoskonale. Sztuczne Państwo przekracza swoją konstytucyjną granicę w momencie, gdy przestaje projektować system dla obywatela mającego samodzielnie zdecydować, a zaczyna projektować obywatela pod wynik, który system ma osiągnąć. Kluczowe nie jest zatem pytanie, czy AI może wpływać na wyborcę.

Ludzie oddziaływali na siebie od zarania polityki. Pytanie przełomowe brzmi inaczej: czy demokratyczne społeczeństwo dopuści do powstania infrastruktury, która zna każdego obywatela indywidualnie, potrafi rozmawiać z każdym z osobna, testować różne strategie perswazji, nigdy się nie męczy i pozostaje własnością podmiotu zainteresowanego konkretnym wynikiem politycznym? Jeśli odpowiedź brzmi tak, problem Sztucznego Państwa przestaje dotyczyć wyłącznie tego, kto liczy głosy. Dotyczy tego, kto współtworzy warunki, w których człowiek decyduje, jaki głos oddać.

Następnym krokiem nie może być wezwanie do odrzucenia maszyn. Musi nim być pytanie konstytucyjne: jak podporządkować predykcję i generatywną perswazję prawom obywatela, zamiast podporządkowywać obywatela logice predykcji? Odpowiedź prowadzi ku europejskim próbom konstytucjonalizacji władzy cyfrowej – ochronie danych, regulacji platform, prawu

konkurencji oraz AI Act – i do rozstrzygnięcia, czy demokratyczne państwo wciąż dysponuje instrumentami zdolnymi objąć regułami infrastrukturę, która coraz skuteczniej przewiduje zachowania jego obywateli. Konstytucjonalizacja algorytmu.

Europa wobec Sztucznego Państwa: od ochrony danych do prawa do ludzkiego osądu. Sztuczne Państwo powstaje wtedy, gdy zdolność techniczna zaczyna zastępować polityczną legitymizację. Jego demontaż nie musi jednak oznaczać niszczenia infrastruktury cyfrowej. Bardziej adekwatnym pojęciem jest konstytucjonalizacja algorytmu: podporządkowanie systemów obliczeniowych normom określającym, czego władza – publiczna lub prywatna – może dowiedzieć się o człowieku, jak może tę wiedzę wykorzystać, kiedy wolno jej automatyzować decyzje, jakie obowiązki przejrzystości musi spełnić oraz w jaki sposób człowiek może tę decyzję zakwestionować. Konstytucjonalizacja nie oznacza tutaj literalnego wpisania algorytmu do konstytucji, lecz przeniesienie zasad państwa prawa – legalności, proporcjonalności, odpowiedzialności, kontroli, praw jednostki i rozdzielenia władz – do sfery, w której realną władzę sprawują systemy techniczne i ich prywatni właściciele.

Europejska odpowiedź na cyfrową koncentrację władzy jest pod tym względem interesująca, ponieważ nie powstała jako jeden, monolityczny akt regulujący „technologię". Jest raczej warstwowym systemem norm, z których każda uderza w inne źródło tej władzy: RODO ogranicza możliwości wykorzystywania danych osobowych i profilowania; Digital Services Act reguluje odpowiedzialność oraz ryzyka środowiska platformowego; Digital Markets Act koncentruje się na gospodarczej pozycji cyfrowych gatekeeperów, natomiast AI Act buduje system wymagań zależnych od rodzaju i poziomu ryzyka stwarzanego przez konkretny system sztucznej inteligencji.

Ochrona podmiotowości poprzez RODO i DSA

Wspólnie tworzą one strukturę, którą można nazwać europejską konstytucją infrastruktury cyfrowej, choć żaden z tych aktów samodzielnie taką funkcję nie pełni. Fundamentem tej konstrukcji pozostaje ochrona danych osobowych, której znaczenie wykracza daleko poza prywatność rozumianą jako prawo do ukrywania informacji.

W społeczeństwie predykcyjnym stawką jest bowiem możliwość wytwarzania wiedzy o człowieku i przekształcania jej w decyzję. RODO adresuje ten problem m.in. poprzez regulację profilowania i zautomatyzowanego podejmowania decyzji. Komisja Europejska wyjaśnia, że profilowanie obejmuje ocenianie osobistych cech człowieka w celu formułowania przewidywań dotyczących jego zachowania lub właściwości; jednocześnie osoba ma co do zasady prawo nie podlegać decyzji opartej wyłącznie na automatycznym przetwarzaniu, jeżeli wywołuje ona skutki prawne albo w

podobnie istotny sposób na nią wpływa. W sytuacjach, gdy taka decyzja jest dopuszczalna, muszą istnieć zabezpieczenia obejmujące możliwość ludzkiej interwencji, przedstawienia własnego stanowiska i zakwestionowania rozstrzygnięcia. Jest to rozwiązanie o znaczeniu filozoficznym znacznie większym niż techniczna regulacja bankowego scoringu. Zawiera ono założenie antropologiczne: człowiek nie powinien zostać całkowicie utożsamiony z wynikiem modelu, który go opisuje. Nawet jeśli statystyczna predykcja jest trafna, osoba zachowuje prawo do przedstawienia indywidualnego przypadku. Powracamy więc do rozróżnienia ustanowionego wcześniej w genealogii statystyki: prawdopodobieństwo populacyjne nie jest jeszcze pełną prawdą o jednostce.

Prawo do ludzkiej interwencji jest instytucjonalnym sposobem przypomnienia systemowi, że mapa nie jest terytorium. Nie można jednak fetyszyzować samej obecności człowieka. Jeżeli urzędnik albo pracownik banku jedynie zatwierdza wynik algorytmu bez realnej zdolności jego zakwestionowania, "human in the loop" staje się dekoracją. Psychologia automatyzacji zna zjawisko automation bias: ludzie mogą nadmiernie ufać rekomendacji systemu właśnie dlatego, że jest ona przedstawiana jako rezultat technicznej analizy. Konstytucjonalizacja algorytmu musi więc wymagać czegoś silniejszego niż biologiczna obecność człowieka przy ekranie. Potrzebna jest realna zdolność do zmiany wyniku, dostęp do wystarczających informacji oraz warunki organizacyjne pozwalające ponieść odpowiedzialność za własny osąd. "Human oversight" nie powinno być rozumiane jako rytualne kliknięcie "zatwierdź". Prawdziwy nadzór wymaga kompetencji, czasu, autonomii wobec systemu oraz odpowiedzialności. Im bardziej złożony model, tym większe ryzyko, że człowiek stanie się jedynie końcowym interfejsem automatycznej decyzji. To jedna z najbardziej subtelnych dróg do automatokracji: formalnie decyzję nadal podejmuje osoba, lecz poznawczo nie posiada już zdolności do działania niezależnego od systemu.

Druga warstwa europejskiej odpowiedzi, Digital Services Act, przesuwa punkt ciężkości z danych pojedynczej osoby na środowisko informacyjne całego społeczeństwa. Jest to szczególnie istotne w kontekście analizy platform jako systemów selekcji. DSA nie zakłada, że ranking czy rekomendacja są z natury nielegalne, lecz nakłada na największe platformy i wyszukiwarki szczególne obowiązki w zakresie ryzyk systemowych, przejrzystości reklam, dostępu do danych, audytów oraz funkcjonowania systemów rekomendacyjnych. Bardzo duże platformy muszą między innymi oferować użytkownikowi co najmniej jedną opcję systemu rekomendacji, która nie opiera się na profilowaniu, oraz prowadzić publicznie dostępne repozytoria reklam.

W tym rozwiązaniu kryje się interesująca teoria wolności. Prawo nie zakazuje personalizacji, lecz odmawia uznania jej za jedyny możliwy sposób organizacji świata informacyjnego. Użytkownik ma otrzymać możliwość wyjścia z części logiki profilowania bez konieczności całkowitego opuszczania

platformy. Jest to subtelna próba przekształcenia architektury cyfrowej z systemu "take it or leave it" w środowisko, w którym jednostka zachowuje wybór dotyczący sposobu, w jaki algorytm pośredniczy w jej kontakcie ze światem. DSA jest tym bardziej istotny, że traktuje system rekomendacyjny nie tylko jako prywatną funkcjonalność produktu, lecz jako potencjalne źródło ryzyka systemowego. Komisja badała m.in., w jaki sposób konstrukcja rekomendacji może zagrażać procesom wyborczym, dyskursowi obywatelskiemu, zdrowiu psychicznemu, ochronie małoletnich oraz rozpowszechnianiu nielegalnych treści. To zasadnicza zmiana myślenia o platformie: algorytm rankingowy przestaje być wyłącznie tajemnicą handlową przedsiębiorstwa, a jego społeczny efekt staje się przedmiotem publicznego nadzoru, co jest szczególnie widoczne w odniesieniu do wyborów.

Rozbijanie monopolu informacyjnego poprzez regulacje DSA i DMA

Komisja Europejska przyjęła wytyczne dotyczące ograniczania systemowych ryzyk wyborczych przez bardzo duże platformy i wyszukiwarki. Wskazała w nich m.in. na systemy rekomendacyjne, transparentność reklam, niezależne audyty oraz dostęp badaczy do danych jako kluczowe elementy instrumentarium prawnego. W 2024 roku wszczęto postępowanie wobec TikToka w związku z podejrzeniem niewystarczającego zarządzania ryzykiem dla integralności wyborów w Rumunii, ze szczególnym uwzględnieniem algorytmów rekomendacyjnych, skoordynowanej manipulacji oraz płatnych treści politycznych.

Z perspektywy wcześniejszej analizy „produkcji preferencji” jest to zmiana o fundamentalnym znaczeniu. Prawo nie próbuje rozstrzygnąć, czy algorytm faktycznie „zmienił wynik wyborów”, co empirycznie często byłoby niemożliwe. Pyta raczej, czy platforma stworzyła i odpowiednio zarządzała ryzykiem dla warunków demokratycznej konkurencji. Przedmiotem regulacji staje się zatem infrastruktura procesu, a nie tylko jego finalny skutek. W 2026 roku podejście to zaczęło obejmować bezpośrednio generatywną AI – Komisja prowadziła wówczas postępowania dotyczące integracji Groka z platformą X i wynikających z niej ryzyk systemowych.

Dnia 31 sierpnia 2026 roku Komisja ogłosiła, że ChatGPT został wyznaczony jako Very Large Online Search Engine, natomiast Reddit i Roblox jako Very Large Online Platforms w rozumieniu DSA. To wydarzenie ma charakter symboliczny: granica między wyszukiwarką, platformą a generatywnym pośrednikiem informacji zaczyna się prawnie zacierać, gdyż system konwersacyjny może pełnić funkcję infrastruktury dostępu do wiedzy.

Trzecia europejska warstwa regulacyjna dotyczy nie tyle treści, co struktury rynku. Digital Markets Act (DMA) wychodzi z założenia, że pewne platformy osiągają pozycję gatekeeperów – bram, przez które inne przedsiębiorstwa i użytkownicy muszą przechodzić, aby uczestniczyć w gospodarce cyfrowej. DMA nie zastępuje klasycznego prawa konkurencji, lecz uzupełnia je zestawem obowiązków i zakazów ex ante. Gatekeeperzy muszą między innymi umożliwiać określone formy interoperacyjności, udostępniać użytkownikom biznesowym dane wytworzone w toku korzystania z platformy oraz nie mogą uprzywilejwiwać własnych usług w rankingach względem ofert konkurencji.

Rozwiązanie to można odczytać jako ekonomiczny odpowiednik zasady podziału władz. Problem Sztucznego Państwa polega m.in. na kumulowaniu wielu funkcji w jednym podmiocie: właściciel infrastruktury może być jednocześnie uczestnikiem rynku, twórcą reguł dostępu, sędzią sporu i beneficjentem rankingu. DMA próbuje rozszczelnić tę koncentrację. Nie zakazuje posiadania platformy, lecz ogranicza możliwość wykorzystania pozycji administratora do arbitralnego wzmacniania własnej działalności.

Interoperacyjność ma znaczenie polityczne znacznie większe niż tylko techniczna wygoda. Zamknięty ekosystem drastycznie zwiększa koszt wyjścia. Jeśli użytkownik nie może łatwo przenieść danych, kontaktów czy relacji, formalna możliwość opuszczenia platformy staje się iluzoryczna. Interoperacyjność redukuje władzę wynikającą z efektu zamknięcia i przekształca Hirschmanowskie „exit” z abstrakcyjnego prawa w realną możliwość. W społeczeństwie platformowym może ona zatem pełnić funkcję analogiczną do prawa do zmiany dostawcy infrastruktury bez utraty całej społecznej tożsamości.

Tę logikę w coraz większym stopniu dotyka również problem chmury i AI. Komisja prowadzi dyskusję nad pozycją największych dostawców infrastruktury chmurowej, a DMA posiada już mechanizmy pozwalające analizować nowe formy gatekeeping power. Jest to szczególnie istotne, ponieważ konkurencja w obszarze AI nie rozgrywa się wyłącznie na poziomie aplikacji. Jeśli dostęp do procesorów, chmury i modeli bazowych pozostanie skoncentrowany w rękach kilku przedsiębiorstw, otwarty rynek aplikacji może współistnieć z oligopolistyczną strukturą infrastrukturalną.

AI Act jako próba regulacji asymetrii władzy algorytmicznej

Najbardziej bezpośrednią europejską próbą uregulowania sztucznej inteligencji jest AI Act. Jego znaczenie dla teorii Sztucznego Państwa przejawia się przede wszystkim w przyjęciu zasady, że nie wszystkie zastosowania AI są politycznie równoważne. System poprawiający zdjęcie, algorytm rekrutacyjny, system identyfikacji biometrycznej czy model wspierający decyzje migracyjne nie powinny podlegać identycznemu reżimowi tylko dlatego, że wszystkie określa się mianem „AI". Władza systemu wynika bowiem z kontekstu jego użycia, skali możliwych naruszeń praw oraz asymetrii między podmiotem wdrażającym technologię a osobą poddaną jej działaniu.

Stan prawny na 2 września 2026 roku jest jednak bardziej złożony niż proste stwierdzenie, że AI Act już w pełni obowiązuje. Rozporządzenie weszło w życie 1 sierpnia 2024 roku i zasadniczo stało się stosowane 2 sierpnia 2026 roku. Zakazane praktyki oraz pierwsze obowiązki zaczęły obowiązywać już od 2 lutego 2025 roku, natomiast przepisy dotyczące zarządzania i modeli ogólnego przeznaczenia weszły w zastosowanie 2 sierpnia 2025 roku. W wyniku AI Omnibus, który wszedł w życie 27 lipca 2026 roku, terminy stosowania najważniejszych wymogów dla części systemów wysokiego ryzyka zostały jednak przesunięte: dla zastosowań z Annex III – obejmujących między innymi biometrię, edukację, zatrudnienie, infrastrukturę krytyczną, migrację i kontrolę graniczną – do 2 grudnia 2027 roku, a dla systemów AI wbudowanych w produkty regulowane do 2 sierpnia 2028 roku.

To przesunięcie ma podwójne znaczenie. Z jednej strony obrazuje trudność w przełożeniu wysokich aspiracji prawnych na konkretne standardy techniczne, procedury oceny zgodności i realną zdolność administracyjną. Z drugiej zaś ujawnia klasyczne napięcie Sztucznego Państwa: technologia rozwija się w czasie przemysłowym, natomiast demokracja reguluje ją w czasie proceduralnym. Jeśli wymagania prawne wchodzą w życie kilka lat po rozpoczęciu masowych wdrożeń, część architektury społecznej może zostać już trwale ustalona. Prawo ponownie staje przed problemem path dependence. Nie oznacza to jednak, że AI Act pozostaje w 2026 roku wyłącznie obietnicą przyszłości.

Jednym z najbardziej znaczących obowiązujących elementów stał się artykuł 50. Od 2 sierpnia 2026 roku nakłada on określone obowiązki przejrzystości dotyczące systemów interaktywnych i generatywnych. Człowiek powinien zostać poinformowany, gdy wchodzi w bezpośrednią interakcję z systemem AI; dostawcy muszą w określonych przypadkach umożliwiać maszynowe wykrywanie treści wygenerowanych lub zmanipulowanych przez AI. Odrębne obowiązki dotyczą deepfake'ów, systemów rozpoznawania emocji, kategoryzacji biometrycznej oraz publikowania generowanych

przez AI tekstów dotyczących spraw interesu publicznego bez odpowiedniej kontroli ludzkiej. Prawny sens tych wymogów jest łatwy do przeoczenia, jeśli sprowadzi się je jedynie do etykiety „AI-generated". W istocie artykuł 50 tworzy zalążek prawa do wiedzy, z jakiego rodzaju podmiotem komunikacyjnym mamy do czynienia. W epoce opisanej w poprzedniej części jest to warunek autonomii.

Jeżeli agent może prowadzić perswazyjną rozmowę, podstawową informacją konstytutywną dla relacji powinno być to, czy rozmówcą jest człowiek, maszyna albo hybrydowy system działający w imieniu określonej organizacji. Sama transparentność nie rozwiązuje oczywiście problemu manipulacji – użytkownik może mieć świadomość, że rozmawia z AI i mimo to ulegać bardzo skutecznej perswazji.

AI Act jako granica normatywna optymalizacji

Etykieta nie neutralizuje asymetrii. Jest jednak warunkiem wcześniejszym: bez identyfikacji podmiotu nie można nawet rozpocząć świadomej oceny jego intencji i interesu. AI Act idzie dalej, uznając część praktyk nie za ryzykowne, lecz za niedopuszczalne. Artykuł 5 zakazuje określonych zastosowań sprzecznych z europejskimi wartościami, w tym pewnych form szkodliwej manipulacji i wykorzystywania podatności, social scoringu, wybranych odmian predykcyjnego profilowania policyjnego, rozpoznawania emocji w miejscu pracy i edukacji, niektórych kategorii klasyfikacji biometrycznej oraz – z ograniczonymi wyjątkami – zdalnej identyfikacji biometrycznej w czasie rzeczywistym do celów egzekwowania prawa w przestrzeni publicznej. Komisja wprost wiąże te zakazy z ochroną godności, wolności, równości, demokracji, praworządności i praw podstawowych. W tym miejscu europejski model wykonuje ruch o charakterze konstytucyjnym: stwierdza, że nie wszystko, co można zoptymalizować, wolno optymalizować. Rynek nie otrzymuje pełnej przestrzeni możliwości technicznych; pewne rozwiązania są z niej usuwane z powodów normatywnych. Jest to dokładne przeciwieństwo radykalnego determinizmu technologicznego. Fakt, że system potrafi wykonać daną operację, nie tworzy prawa do jej wykonywania.

W przypadku systemów wysokiego ryzyka logika jest odmienna. Nie zakazuje się ich z definicji, lecz ustanawia strukturę odpowiedzialności, dokumentacji, kontroli jakości, zarządzania ryzykiem, monitorowania i nadzoru ludzkiego. Szczególnie istotny dla teorii politycznej jest obowiązek oceny wpływu na prawa podstawowe w określonych zastosowaniach. Publiczne organy i podmioty świadczące usługi publiczne, wykorzystujące objęte przepisami systemy wysokiego ryzyka, mają oceniać wpływ ich wdrożenia na fundamentalne prawa przed pierwszym użyciem; systemy wspierające decyzje dotyczące osób podlegają również obowiązkom informacyjnym.

Klasyczna ocena pytania „czy algorytm działa?” zostaje zatem rozszerzona o pytanie: „co jego działanie robi z człowiekiem posiadającym prawa?”. To zmiana języka. Inżynier może zmierzyć accuracy, precision, recall, false-positive rate albo latency. Konstytucjonalista musi natomiast zapytać o dyskryminację, prawo do obrony, autonomię, prywatność, wolność wypowiedzi i możliwość odwołania. Te dwa porządki wiedzy nie są konkurencyjne, lecz komplementarne, ponieważ technicznie doskonały system może nadal pozostawać konstytucyjnie niedopuszczalny. Dobrym przykładem jest algorytm zdolny bardzo trafnie przewidywać ryzyko popełnienia określonego czynu. Z technicznego punktu widzenia poprawa predykcji wydaje się jednoznacznie pożądana. Z prawnego zaś pojawia się pytanie, czy państwo może ingerować w prawa konkretnej osoby na podstawie statystycznego prawdopodobieństwa zachowania, którego jeszcze nie podjęła. Im większa dokładność modelu, tym silniejsza pokusa rozszerzenia jego stosowania. Konstytucjonalizm musi więc stawiać granice właśnie wtedy, gdy technologia staje się skuteczna, a nie tylko wtedy, gdy popełnia błędy.

Z tej perspektywy AI Act, RODO, DSA i DMA można odczytać jako cztery różne odpowiedzi na cztery wymiary Sztucznego Państwa. RODO mówi, że człowiek nie jest bezpańskim zbiorem danych, który można dowolnie profilować. DSA wskazuje, że środowisko komunikacyjne nie jest czysto prywatnym interfejsem, jeśli jego architektura wpływa na całe społeczeństwa. DMA mówi, że właściciel infrastrukturalnej bramy nie może korzystać z pełnej swobody zwykłego uczestnika rynku. AI Act mówi zaś, że moc predykcyjna i automatyzacyjna systemu musi pozostawać proporcjonalna do prawnego i społecznego ryzyka jego zastosowania.

Nie jest to jeszcze pełny demontaż Sztucznego Państwa, lecz próba fragmentacji jego władzy. Ta fragmentacja ma znaczenie analogiczne do klasycznego podziału władz. Monteskiusz nie zakładał, że człowiek sprawujący władzę stanie się moralnie doskonały; zakładał raczej, że władzę należy powstrzymywać przez władzę. W środowisku cyfrowym podobną funkcję mogą pełnić konkurencja, interoperacyjność, audyt, organy ochrony danych, badacze posiadający dostęp do platformowych danych, sądy, regulatorzy rynku, obowiązki dokumentacyjne i prawa użytkownika.

Ryzyka regulacyjne i konieczność suwerenności epistemicznej państwa

Celem nie jest znalezienie idealnego właściciela algorytmu, lecz stworzenie systemu, w którym nikt nie zyska całkowicie niekontrolowanego pola działania.

Europejski eksperyment spotyka się z poważną krytyką. Nadmierna regulacja może bowiem zwiększać koszty wejścia i utrwalać pozycję największych graczy, dysponujących armią prawników oraz specjalistów od compliance, co w efekcie obniży konkurencyjność europejskich firm względem podmiotów z USA czy Chin. Im bardziej skomplikowany reżim regulacyjny, tym łatwiej poradzi sobie z nim korporacja o miliardowych zasobach niż mały innowator. Regulacja projektowana przeciw koncentracji władzy może zatem paradoksalnie ją umacniać, jeśli koszt zgodności stanie się zbyt wysoką barierą wejścia.

Krytyka ta nie powinna być odrzucana jako lobbyistyczna propaganda; to klasyczny problem ekonomii regulacji. Każda norma generuje koszt zgodności, a koszty stałe obciążają mniejsze przedsiębiorstwa proporcjonalnie mocniej. Właśnie dlatego europejskie prawo stara się stosować zasadę proporcjonalności, różnicując obowiązki w zależności od skali i ryzyka oraz wprowadzając uproszczenia dla mniejszych podmiotów. AI Omnibus z 2026 roku dowodzi, że pierwotne ambicje regulacyjne muszą być konfrontowane z realną wykonalnością, dostępnością standardów i zdolnością rynku do wdrożenia wymogów.

Istnieje jednak również ryzyko regulacyjnej iluzji. Państwo może uchwalić imponujące prawo, pozostając jednocześnie poznawczo słabszym od podmiotów, które reguluje. Audyt jest wart tyle, ile kompetencje audytora; obowiązek dokumentacji pomaga tylko wtedy, gdy regulator potrafi tę dokumentację rzetelnie ocenić, a kara odstrasza jedynie przy wystarczającym prawdopodobieństwie wykrycia naruszenia.

Konstytucjonalizacja algorytmu wymaga więc nie tylko legislacji, lecz realnej zdolności instytucjonalnej państwa: kadr technicznych, laboratoriów, dostępu do mocy obliczeniowej i danych, wsparcia niezależnych badaczy oraz dobrze finansowanych organów nadzoru. W przeciwnym razie może powstać osobliwa odmiana Sztucznego Państwa praworządnego na papierze. Prywatny system będzie technicznie dominował, a administracja ograniczy się do odgrywania rytuał kontroli, nie posiadając zdolności do jego realnego wykonania. Formalna suwerenność prawa nie wystarczy, jeśli państwo utraci suwerenność epistemiczną.

Kluczowa różnica między modelem europejskim a radykalnym techno-libertarianizmem dotyczy zatem nie stosunku do innowacji, lecz pytania o pierwszeństwo normatywne. Techno-libertarianin zakłada, że innowacja powinna być domyślnie dozwolona, a państwo musi udowodnić zasadność jej ograniczenia. Europejski konstytucjonalizm odpowiada: im większa asymetria władzy i potencjalny wpływ na prawa podstawowe, tym silniejsze uzasadnienie dla publicznego wyznaczania granic. Nie jest to konflikt technologii z antytechnologią.

Prawo do niewyredukowanej podmiotowości jako granica automatokracji

Jest to spór o tytuł do ustanawiania funkcji celu.

Gdy przedsiębiorstwo buduje model do generowania obrazów na życzenie użytkownika, znaczna autonomia projektanta jest racjonalna. Jeśli jednak ten sam model lub jego odmiana ocenia, kto otrzyma pracę, kredyt, świadczenie publiczne albo możliwość przekroczenia granicy, społeczny kontekst zmienia naturę technologii. Kod przestaje być wtedy tylko produktem; staje się elementem procedury przydzielającej dobra, prawa i ryzyka.

Zasadniczym kryterium przyszłego konstytucjonalizmu AI nie powinno być zatem pytanie o to, czy decyzję podjęła maszyna, lecz jaki rodzaj władzy został za jej pomocą wykonany. Komputer wspomagający lekarza różni się od komputera automatycznie odmawiającego świadczenia; model doradzający sędziemu różni się od modelu, którego wynik staje się de facto wyrokiem; system pomagający wyborcy zrozumieć program różni się od systemu optymalizowanego tak, aby zmienić jego głos. Przyszłość demokratycznego państwa zależy nie od zachowania „człowieka w każdej pętli", lecz od ustalenia, których pętli nie wolno zamknąć bez człowieka posiadającego rzeczywistą władzę zakwestionowania systemu. Jest to subtelna różnica wobec programu całkowitej automatyzacji.

Automatokracja pyta, gdzie człowieka można jeszcze usunąć z procesu. Konstytucjonalizm powinien natomiast pytać, gdzie obecność ludzkiego osądu jest elementem samej legitymizacji procesu. Sędzia nie jest potrzebny wyłącznie dlatego, że algorytm nie osiągnął jeszcze odpowiedniej dokładności; jest niezbędny, ponieważ wymiar sprawiedliwości obejmuje interpretację, wysłuchanie, odpowiedzialność i publiczne uzasadnienie. Obywatel nie głosuje dlatego, że państwo nie potrafi jeszcze przewidzieć jego wyboru, lecz dlatego, że akt oddania głosu jest częścią jego statusu politycznego. Podobnie lekarz nie powinien być traktowany jako tymczasowy interfejs między pacjentem a przyszłym, doskonałym modelem diagnostycznym, o ile relacja medyczna obejmuje wartości, których sama predykcja nie wyczerpuje.

W tym miejscu pojawia się pojęcie, które może stanowić centralną normatywną kategorię całego artykułu: prawo do niewyredukowanej podmiotowości. Nie oznacza ono prawa do świata wolnego od danych, statystyki i modeli – byłoby to niewykonalne i społecznie szkodliwe. Oznacza ono prawo człowieka do tego, aby w sprawach istotnych dla jego praw, wolności i politycznego statusu nie traktowano modelu statystycznego jako kompletnego odpowiednika osoby. Tak rozumiana

konstytucjonalizacja algorytmu nie walczy więc z maszyną, lecz z pomyleniem reprezentacji człowieka z człowiekiem.

Europejski model pozostaje eksperymentem niedokończonym i może ponieść porażkę. Regulacje mogą okazać się zbyt wolne, zbyt skomplikowane albo nieskutecznie egzekwowane; mogą generować niezamierzone koszty, które osłabią konkurencję i zwiększą przewagę największych podmiotów. Technologie mogą również ewoluować szybciej niż kategorie prawne. Jednak sama próba ich uregulowania jest argumentem przeciwko technologicznej nieuchronności Sztucznego Państwa. Kod może podlegać prawu, model audytowi, a platforma obowiązkom publicznym. Gatekeeper może zostać zobowiązany do otwarcia części infrastruktury, automatyczna decyzja – zakwestionowana, a system perswazyjny – zmuszony do ujawnienia swojej natury. Niektóre technicznie możliwe praktyki mogą zostać po prostu uznane za niedopuszczalne. Oznacza to demontaż politycznego roszczenia maszyny do suwerenności. Nie polega on na wyłączeniu serwerów, lecz na odebraniu technicznej skuteczności prawa do samolegitymizacji.

Pozostaje pytanie najtrudniejsze: co zrobić, jeżeli maszyny rzeczywiście okażą się lepsze od ludzi w wielu zadaniach wykonywanych dziś przez państwo? Jeśli będą dokładniej diagnozować, wykrywać oszustwa, tłumaczyć prawo, alokować zasoby czy prognozować kryzysy, znajdując rozwiązania niewidoczne dla ludzkiego umysłu – czy konsekwentna obrona „ludzkiej dyskrecji" nie stanie się wtedy obroną prawa do gorszej decyzji? Pytanie to musi zostać potraktowane z pełną powagą.

Krytyka Sztucznego Państwa będzie niewiarygodna, jeśli nie skonfrontuje się z najsilniejszym argumentem jego zwolenników: maszyny mogą po prostu rządzić niektórymi fragmentami rzeczywistości lepiej od nas. Kolejny etap wymaga świadomego odwrócenia perspektywy. Zamiast pytać, czego należy się bać w automatyzacji państwa, trzeba zapytać, czego demokratyczne państwo może od niej potrzebować. Dopiero po zbudowaniu najsilniejszej możliwej obrony algorytmicznego rządzenia będzie można wyznaczyć granicę, po której pomoc maszyny przekształca się w automatokrację.

> Czy w świecie, w którym maszyny mogą diagnozować i zarządzać rzeczywistością sprawniej niż jakikolwiek urzędnik czy polityk, uparta obrona ludzkiego osądu stanie się jedynie walką o prawo do popełniania gorszych decyzji? Być może największym wyzwaniem dla współczesnej demokracji nie jest sama obecność algorytmów, lecz konieczność zdefiniowania, które fragmenty naszego człowieczeństwa są zbyt cenne, by pozwolić na ich optymalizację. Prawdziwa wolność może wkrótce oznaczać prawo do bycia mniej efektywnym niż model statystyczny.

Inspiracje

[1] "The Rise and Fall of the Artificial State" Jill Lepore

Jill Lepore (ur. 27 sierpnia 1966) jest cenioną amerykańską historyczką, pisarką i dziennikarką. Jest profesorem historii amerykańskiej im. Davida Woodsa Kempera (rocznik 1941) na Uniwersytecie Harvarda oraz profesorem prawa na Harvard Law School, a od 2005 roku jest również stałym felietonistą „The New Yorker". Lepore uzyskała tytuł doktora amerykanistyki na Uniwersytecie Yale w 1995 roku. Jej badania naukowe koncentrują się na wczesnej historii Ameryki, konstytucjonalizmie, kulturze politycznej oraz powiązaniach technologii i demokracji. Jako powszechnie uznana intelektualistka, Lepore jest ceniona za prozę narracyjną i skrupulatne badania archiwalne. Do jej przełomowych osiągnięć należą nagrodzona nagrodą Bancrofta monografia „The Name of War", bestsellerowe, obszerne studium „These Truths: A History of the United States" oraz uznane przez krytyków książki historyczne analizujące amerykański rozwój polityczny, systemy prawne i instytucje demokratyczne.

Jill Lepore *(born August 27, 1966) is an acclaimed American historian, author, and journalist. She serves as the David Woods Kemper '41 Professor of American History at Harvard University and Professor of Law at Harvard Law School, while also contributing as a staff writer for The New Yorker since 2005. Lepore earned her Ph.D. in American Studies from Yale University in 1995. Her scholarship specializes in early American history, constitutionalism, political culture, and the intersection of technology and democracy. A widely recognized public intellectual, Lepore is celebrated for her narrative prose and rigorous archival research. Her seminal contributions include her Bancroft Prize-winning monograph The Name of War, the bestselling comprehensive survey These Truths: A History of the United States, and critically acclaimed histories examining American political development, legal systems, and democratic institutions.*

Harvard University

Literatura uzupełniająca

Książki

Autorzy przywoływani w artykule:



[1] "If Then"

Jill Lepore

2020 | Hachette UK | ISBN: 9781529386189



[2] "Blindspot"

Jane Kamensky, Jill Lepore

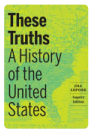
2009 | Random House | ISBN: 9780385526203



[3] "The Mansion of Happiness"

Jill Lepore

2012 | Vintage | ISBN: 9780307958501



[4] "These Truths"

Jill Lepore

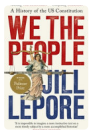
2023 | ISBN: 9781324043799



[5] "The Story of America"

Jill Lepore

2012 | Princeton University Press | ISBN: 9780691159591

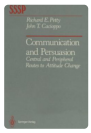


[6] "We the People"

Jill Lepore

2026 | John Murray Publishers | ISBN: 9781399827089

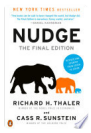
Literatura pokrewna (Google Scholar):



[7] "Communication and Persuasion"

Richard Petty

Springer | ISBN: 9781461249658



[8] "Nudge"

Richard Thaler

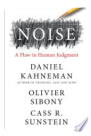
Penguin | ISBN: 9780525508526



[9] "Misbehaving"

Richard Thaler

W. W. Norton & Company | ISBN: 9780393246773



[10] "Noise: A Flaw in Human Judgment"

Cass Sunstein

Hachette UK | ISBN: 9780316451383

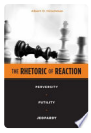
[11] "Changing Conceptions of Administration"

Cass Sunstein

[12] **"Crowdsourcing interventions to promote uptake of COVID-19 booster vaccines"**
Cass Sunstein



[13] **"Exit, Voice, and Loyalty"**
Albert O. Hirschman
Harvard University Press | ISBN: 9780674276604



[14] **"The Rhetoric of Reaction"**
Albert O. Hirschman
Harvard University Press | ISBN: 9780674768680

Artykuły naukowe

Autorzy przywoływani:

[1] **"The European Citizens' Initiative: A Victory for Democracy or a Marketing Trick?"**
Rafał Trzaskowski
2010 | European View | DOI: 10.1007/s12290-010-0133-3
Google Scholar

[2] **"The Elaboration Likelihood Model of Persuasion"**
Richard E. Petty, John T. Cacioppo
1986 | Advances in experimental social psychology | DOI: 10.1016/s0065-2601(08)60214-2
Google Scholar

[3] **"Communication and Persuasion"**
Richard E. Petty, John T. Cacioppo
1986 | DOI: 10.1007/978-1-4612-4964-1
Google Scholar

[4] **"Attitudes and Persuasion: Classic and Contemporary Approaches"**
Richard E. Petty, John T. Cacioppo
1981 | Medical Entomology and Zoology
Google Scholar

[5] **"Personal involvement as a determinant of argument-based persuasion."**
Richard E. Petty, John T. Cacioppo, Rachel E. Goldman
1981 | Journal of Personality and Social Psychology | DOI: 10.1037/0022-3514.41.5.847
Google Scholar

[6] **"From Cashews to Nudges: The Evolution of Behavioral Economics"**

R. Thaler

2018 | The American Economic Review | DOI: 10.1257/aer.108.6.1265

Google Scholar

[7] **"Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations, and States."**

Gordon Tullock, Albert O.HG Hirschman

1970 | The Journal of Finance | DOI: 10.2307/2325604

Google Scholar

[8] **"The Passions and the Interests: Political Arguments for Capitalism before Its Triumph."**

Andrew Stewart Skinner, Albert O.HG Hirschman

1979 | Economica | DOI: 10.2307/2553103

Google Scholar

[9] **"National Power and the Structure of Foreign Trade"**

Henry Oliver, Albert O.HG Hirschman

1946 | Southern Economic Journal | DOI: 10.2307/1052282

Google Scholar

[10] **"The Political Economy of Import-Substituting Industrialization in Latin America"**

Albert O.HG Hirschman

1968 | The Quarterly Journal of Economics | DOI: 10.2307/1882243

Google Scholar

[11] **"Reflections on the causes of the rise and fall of the Roman empire"**

Charles de Secondat Montesquieu

2008

Google Scholar

Artykuły pokrewne (Google Scholar):

[12] **"We the people eine Geschichte der amerikanischen Verfassung"**

Jill 1966- Lepore, Werner Roller, Annabel 1976- Zettel, Verlag Ch Beck

Google Scholar

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: ac435c96-e5cb-4ed1-8ba6-1f53bf3a27c2