Sztuczna inteligencja i sprawiedliwość w służbie demokracji deliberatywnej: od skalowania namysłu po reprezentację przyszłych pokoleń w kontekście myśli Jamesa S. Fishkina.

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje wykorzystanie sztucznej inteligencji do skalowania procesów demokratycznych w duchu myśli Jamesa S. Fishkina. Autor skupia się na przejściu od małych grup obywatelskich do masowych platform deliberatywnych, takich jak Stanford Online Deliberation Platform. Głównym problemem jest kwestia władzy proceduralnej – przeniesienie moderacji na poziom kodu może zamrozić normy polityczne i stworzyć nową formę asymetrii władzy. Tekst zestawia model Habermasowski, kładący nacisk na komunikacyjną legitymację, z proceduralizmem Fishkina, gdzie AI służy redukcji kosztów organizacyjnych. Wnioski wskazują, że choć technologia wspiera dostęp do głosu i logistykę, nie może stać się epistemicznym arbitrem konsensusu, gdyż neutralność algorytmiczna jest mitem, a kod zawsze odzwierciedla cele swoich twórców.

The article analyzes the use of artificial intelligence to scale democratic processes in the spirit of James S. Fishkin's thought. The author focuses on the transition from small citizen groups to mass deliberative platforms, such as the Stanford Online Deliberation Platform. The main issue is the question of procedural power – transferring moderation to the level of code may freeze political norms and create a new form of power asymmetry. The text contrasts the Habermasian model, emphasizing communicative legitimacy, with Fishkin's proceduralism, where AI serves to reduce organizational costs. The conclusions indicate that while technology supports access to

voice and logistics, it cannot become an epistemic arbiter of consensus, as algorithmic neutrality is a myth and code always reflects the goals of its creators.

Artykuł analizuje potencjał i zagrożenia związane z wykorzystaniem sztucznej inteligencji do skalowania demokracji deliberatywnej, na przykładzie sondaży deliberatywnych Jamesa Fishkina. Autor stawia tezę, że choć AI może radykalnie zwiększyć liczbę uczestników debaty publicznej i pomóc w uwzględnieniu interesów grup marginalizowanych oraz przyszłych pokoleń, niesie ona ryzyko zastąpienia realnego sporu politycznego przez algorytmiczny konsensus. Głównym problemem nie jest sama technologia, lecz przeniesienie władzy na poziom kodu – kto projektuje funkcję celu algorytmu, ten de facto projektuje przestrzeń polityczną. Tekst argumentuje, że aby AI wspierała autonomię obywatela, zamiast ją absorbować, musi przejść z modelu „neutralnego moderatora" do transparentnej i audytowalnej architektury, która nie dąży do wygładzenia konfliktów, lecz czyni widzialną strukturę niezgody oraz interesy tych, którzy nie mogą uczestniczyć w procesie (np. przyszłych pokoleń). W ten sposób AI może stać się infrastrukturą sprawiedliwości, o ile pozostanie narzędziem poszerzania pola informacji, a nie autonomicznym arbitrem prawdy i słuszności.

SŁOWA KLUCZOWE

demokracja deliberatywna | sztuczna inteligencja w polityce | skalowanie namysłu | James S. Fishkin | Habermas Machine | neutralność algorytmiczna | konstytucjonalizm deliberacji | sprawiedliwość międzypokoleniowa | twierdzenie Condorceta | epistemiczna teoria demokracji | deliberative polling | redukcja kosztu proceduralnego | asymetria władzy cyfrowej | zasada zachowania opcji przyszłości | infrastruktura poznawcza demokracji

Skalowanie deliberacji przez AI przenosi władzę na poziom kodu

Algorytm przy stole obrad: sztuczna inteligencja, skalowanie deliberacji i problem nowej władzy proceduralnej

Ambitny etap projektu demokracji deliberatywnej rozpoczyna się w chwili, gdy przestajemy pytać, czy kilkuset obywateli potrafi wspólnie myśleć, a zaczynamy zastanawiać się, czy podobne warunki namysłu można stworzyć dla setek tysięcy lub milionów ludzi. W tym punkcie problem ustrojowy zbiega się z informatyką, teorią decyzji, epistemologią społeczną, psychologią poznawczą, teorią informacji oraz etyką sztucznej inteligencji. Fishkinowski eksperyment przestaje być wówczas wyłącznie metodą badania opinii publicznej i zaczyna przypominać propozycję budowy infrastruktury poznawczej demokracji. Kierunek ten wiąże się ze Stanford Online Deliberation Platform oraz automatyczną moderacją, która umożliwia prowadzenie wielu równoległych dyskusji. Jeżeli jednak moderator zostaje częściowo zastąpiony przez kod, pojawia się pytanie konstytucyjne nowego rodzaju: kto kontroluje reguły kodu, który z kolei kontroluje reguły rozmowy?

Problem skalowania jest realny. Klasyczny Deliberation Day Bruce'a Ackermana i Jamesa Fishkina zakładał organizację masowych, fizycznych spotkań obywatelskich przed wyborami. Koncepcja ta była intelektualnie atrakcyjna, lecz logistycznie wręcz gigantyczna – wymagałaby tysięcy miejsc, moderatorów, materiałów oraz niezliczonych równoległych sesji. Współczesna platforma Stanfordu próbuje przenieść część tych kosztów na infrastrukturę cyfrową.

W eksperymencie opisanym w 2025 roku w „Perspectives on Politics” automatyczny system obsługiwał synchroniczne grupy liczące około dziesięciu osób. Zarządzał kolejką wypowiedzi, przydzielał czas na interwencję (obecnie 45 sekund), zachęcał milczących do udziału, reagował na nieuprzejmy język, tworzył niemal natychmiastowe transkrypcje, prowadził grupę przez punkty agendy i pomagał wyłaniać pytania do części plenarnej. Co istotne, autorzy określają to rozwiązanie jako AI-assisted, a nie jako autonomicznego moderatora: w kluczowych momentach grupa sama głosuje, czy temat został wystarczająco omówiony i czy można przejść dalej. Wyłania się stąd obraz AI gwarantującej równowagę i neutralność rozmowy, jednak nauka nie pozwala na taką pewność. Technologia może wspierać równy dostęp do głosu, ale neutralność algorytmiczna nie jest właściwością wynikającą automatycznie z braku ludzkiego moderatora.

Algorytm posiada bowiem funkcję celu, zestaw reguł, dane treningowe, klasyfikatory i progi interwencji. Ten, kto ustala, czym jest „nieuprzejmość”, jak długo wolno mówić, kiedy problem uznaje się za omówiony, jakie wypowiedzi są semantycznie podobne, które pytania trafiają do

plenarium i jak konstruowana jest agenda – ten w istocie uczestniczy w projektowaniu przestrzeni politycznej. Kod nie usuwa normatywności procedury; kod ją zamraża.

W tym kontekście pojawia się pierwszy z kilku konkurencyjnych modeli teoretycznych. Model Habermasowski interpretuje demokratyczną legitymację poprzez komunikację, w której uczestnicy mogą wzajemnie kwestionować roszczenia do prawdy, słuszności i szczerości, a polityczna władza komunikacyjna powstaje dzięki publicznemu obiegowi argumentów. Habermas nie zakłada, że realne społeczeństwa osiągają idealne warunki rozmowy; założenia te służą raczej jako kryterium krytyki istniejących asymetrii władzy. W jego teorii sfera publiczna ma być przestrzenią, w której decyzje pozostają otwarte na racjonalną krytykę, a komunikacyjna władza obywateli zostaje sprzężona z administracyjną władzą państwa. Z tego punktu widzenia AI może być użyteczna, jeśli usuwa przeszkody w komunikacji: tłumaczy języki, umożliwia uczestnictwo osobom nieśmiałym, porządkuje kolejkę, identyfikuje powtarzające się argumenty lub pomaga wydobyć mniejszościową opinię z masy tekstu.

Nie może ona jednak otrzymać epistemicznego statusu arbitra, który wskazuje grupie jej „racjonalny konsensus". Habermasowskie uzasadnienie nie polega bowiem na maksymalizacji podobieństwa stanowisk, lecz na możliwości wzajemnego żądania racji. Jeśli algorytm syntetyzuje stanowiska tak skutecznie, że uczestnicy przestają konfrontować się ze źródłami własnej niezgody, uzyskujemy konsensus semantyczny bez pełnej deliberacji politycznej.

Drugi model można określić jako Fishkinowski proceduralizm eksperymentalny. Jego centralną wartością nie jest osiągnięcie konsensusu, lecz poprawienie warunków formowania indywidualnego sądu. Dlatego Deliberative Polling nie wymaga jednomyślności – ostateczne odpowiedzi pozostają indywidualne i poufne. AI jest w tej perspektywie przede wszystkim technologią redukcji kosztu proceduralnego: ma pozwolić odtwarzać na dużą skalę warunki podobne do tych, jakie zapewnia dobrze moderowane spotkanie twarzą w twarz.

Wyniki eksperymentu „America in One Room: Climate and Energy" są dla tej hipotezy istotne. Do deliberacji online zrekrutowano 962 uczestników (grupa kontrolna liczyła 661 osób); badano 72 propozycje, a w 66 przypadkach uczestnicy istotnie zmienili stanowiska w toku procesu. Wśród najbardziej spolaryzowanych pytań 22 z 27 wykazały znaczące przesunięcie względem grupy kontrolnej. W przypadku republikańskich uczestników początkowe poparcie dla wielu działań klimatycznych oscylowało wokół 35 procent, by po deliberacji w szeregu spraw przekroczyć 55 procent. Badacze odnotowali także zmianę wskaźników związanych z przekonaniami faktograficznymi dotyczącymi klimatu oraz wzrost obaw o jego konsekwencje, przy braku analogicznych zmian w grupie kontrolnej.

Nie należy jednak wyciągać wniosku, że „AI przekonała Republikanów do polityki klimatycznej". Interwencja obejmowała zbilansowane materiały, strukturę grupową, dyskusję, pytania, czas refleksji oraz automatyczną moderację. AI była częścią pakietu przyczynowego, a nie izolowanym eksperymentalnie czynnikiem. Trzeci model pochodzi z epistemicznej teorii demokracji.

Większa grupa zwiększa szansę na poprawność

Twierdzenie ławy przysięgłych Condorceta dowodzi, że w określonych warunkach decyzja większości zwiększa prawdopodobieństwo trafnego rozstrzygnięcia. Jeśli kompetencja indywidualna uczestników jest wyższa niż losowa, a ich błędy pozostają dostatecznie niezależne, szansa na poprawny wynik rośnie wraz z liczebnością grupy.

Ryzyko homogenizacji i różnica między trafnością a legitymacją

Dla sztucznej inteligencji rodzi to fascynujący paradoks. AI może zwiększać kompetencje jednostek, dostarczając informacji i porządkując argumenty, ale jednocześnie może zmniejszać niezależność ich błędów, ponieważ wszyscy otrzymują treści przetworzone przez ten sam model. Jeżeli tysiąc obywateli myli się na tysiąc niezależnych sposobów, agregacja pozwala część tych błędów wygasić. Jeśli jednak korzystają z tego samego systemu AI, który popełnia jeden systematyczny błąd, otrzymujemy tysiąc skorelowanych odpowiedzi. W teorii informacji i statystyce jest to różnica między szumem niezależnym a błędem wspólnym.

Pierwszy można częściowo redukować poprzez agregację; drugi skaluje się wraz z systemem. Slogan „AI zwiększy mądrość tłumu" jest zatem naukowo zbyt uproszczony. Mądrość tłumu zależy bowiem nie tylko od liczby umysłów, ale od różnorodności informacji i niezależności procesu ich przetwarzania. Jeżeli model językowy stanie się centralnym dostawcą podsumowań, faktów czy kontrargumentów, może poprawić średnią kompetencję poznawczą, lecz zarazem homogenizować wejścia. Z punktu widzenia Condorceta i współczesnej epistemologii grupowej należy więc projektować nie jeden „najlepszy umysł sztuczny", lecz system odporny na skorelowany błąd.

Kolejne ograniczenie wprowadza epistemic proceduralism Davida Estlunda. Nawet gdybyśmy potrafili wykazać, że system wspomagany przez AI statystycznie częściej osiąga trafne rezultaty, nie nadaje mu to automatycznie demokratycznej legitymacji. Estlund odrzuca prostą zasadę, według której rządzić powinien ten, kto jest najbardziej kompetentny. Procedura musi być nie tylko

skuteczniejsza niż przypadek, ale i możliwa do uzasadnienia obywatelom z perspektyw uznanych za kwalifikowane. AI o wysokiej trafności, której działania nie można zakwestionować, właściciel nie podlega kontroli, a błędów nie da się audytować, mogłaby być epistemicznie efektywna, a zarazem demokratycznie nieuprawniona. To zasadnicza różnica między accuracy a legitimacy.

Demokracja nie jest konkursem na najlepszy predyktor. Jeżeli system AI przewidziałby z dziewięćdziesięcioprocentową skutecznością, która reforma przyniesie największy wzrost PKB, nie uzyskuje przez to prawa do samodzielnego decydowania o tym, jaka część społeczeństwa zostanie obciążona kosztami tych zmian. Fakt opisuje prawdopodobny skutek; polityka musi natomiast zważyć wartości, dystrybucję ryzyka oraz moralnie dopuszczalne granice działania.

Piąty model dostarcza teoria wyboru społecznego. Arrow wykazał, że przy określonych, pozornie rozsądnych warunkach nie istnieje idealna reguła agregowania indywidualnych preferencji w spójny ranking społeczny. Literatura deliberatywna nie rozwiązuje tego problemu magicznie, choć Dryzek i Christian List argumentują, że deliberacja może zmieniać samą przestrzeń preferencji, pomagając osiągnąć bardziej uporządkowane rozstrzygnięcie kolektywne. Prowadzi to do istotnego wniosku: AI może oddziaływać nie tylko na agregację istniejących preferencji, ale na ich transformację przed tym procesem. To władza znacznie większa.

Algorytm wyborczy zliczający głosy działa na końcu procesu. Algorytm deliberacyjny może działać znacznie wcześniej: streszcza stanowiska, dobiera podobieństwa, formułuje wspólne mianowniki i wpływa na to, jak uczestnik konceptualizuje przedmiot sporu. Jego rola przypomina zatem nie tyle komisję wyborczą, co połączenie moderatora, redaktora, nauczyciela i agenda-settera. W ten sposób z teorii wyboru społecznego przechodzimy do teorii preferencjotwórczej infrastruktury politycznej.

Pułapka konsensusu w algorytmicznej deliberacji

Najbardziej spektakularnym eksperymentem w tej dziedzinie jest opracowana przez badaczy Google DeepMind tzw. Habermas Machine. W badaniach opublikowanych w Science w 2024 roku system LLM tworzył kolejne wersje grupowych stanowisk, opierając się na opiniach i krytyce uczestników. W eksperymentach wzięło udział 5734 osób; wygenerowane przez AI teksty były częściej preferowane niż te przygotowywane przez ludzkich mediatorów, a uczestnicy oceniali je jako bardziej klarowne, informacyjne i bezstronne.

Badanie wykazało zmniejszenie rozbieżności stanowisk oraz zdolność systemu do uwzględniania perspektyw mniejszościowych przy zachowaniu pozycji większości. Wynik ten powtórzono w

wirtualnym zgromadzeniu z demograficznie reprezentatywną próbą brytyjską. Choć rezultat jest naukowo fascynujący, popularna interpretacja mówiąca o tym, że „AI odkryła konsensus", jest zbyt daleko idąca. System został bowiem zaprojektowany właśnie do generowania propozycji uzyskujących wysoki poziom akceptacji – jego sukces mierzymy zatem kryterium wpisanym w funkcję celu. Tu pojawia się kluczowa dla teorii optymalizacji kwestia: czy maksymalizacja aprobaty jest tożsama z maksymalizacją jakości deliberacji? Nie musi być.

Grupa może osiągnąć porozumienie, ponieważ odkryła najlepsze rozwiązanie, ale może też dojść do niego dlatego, że formuła stała się dostatecznie ogólna, łagodna i wieloznaczna, by każdy rozpoznał w niej fragment własnego stanowiska. W dyplomacji nazywalibyśmy to konstruktywną niejednoznacznością. W nauce politycznej należy natomiast pytać, czy taka synteza usuwa rzeczywisty konflikt, czy jedynie tworzy tekst, pod którym wiele osób może się podpisać. Stało się to przedmiotem ostrej krytyki w literaturze z 2025 roku.

Nicolás Palomo Hernández argumentuje, że Habermas Machine opiera się na wąskiej interpretacji deliberacji: nadmiernie uprzywilejowuje zgodę, przenosi część procesu do sfery indywidualnych i prywatnych ocen oraz redukuje ludzkie uczestnictwo do dostarczania i oceniania stanowisk przetwarzanych przez algorytm. Alessandro Volpe wskazuje z kolei na problemy dotyczące natury porozumienia, inkluzywności moralnej i ukrytej racjonalności systemu. Krytyka ta nie obala wyników eksperymentalnych, lecz pokazuje coś innego: empiryczny sukces funkcji celu nie dowodzi normatywnej poprawności samej funkcji celu.

W tym punkcie rozchodzą się drogi Habermasa i „Habermas Machine". W teorii Habermasa orientacja na porozumienie nie oznacza, że każda rozmowa powinna kończyć się konsensusem. Uczestnicy muszą być gotowi poddać swoje stanowiska racjonalnej krytyce; nie mogą natomiast być programowani na zgodę jako obligatoryjny rezultat. Krytycy wskazują, że technologia maksymalizująca akceptowalność wspólnego zdania może przesunąć deliberację od pytania „który argument jest uzasadniony?" ku pytaniu „które zdanie spodoba się największej liczbie osób?". To zupełnie inny problem optymalizacyjny.

Z punktu widzenia teorii demokracji agonistycznej problem jest jeszcze poważniejszy. Chantal Mouffe i inni krytycy deliberatywizmu przypominaliby, że polityczne konflikty często nie wynikają z braku sprawnego podsumowania. Dotyczą one interesów, władzy, własności, redystrybucji i sprzecznych wizji dobra. Pracownik może dążyć do wyższej części wartości dodanej, a właściciel do niższych kosztów pracy; wspólne sformułowanie o „sprawiedliwym i konkurencyjnym systemie wynagradzania" nie usuwa materialnej sprzeczności.

Algorytm znajdujący wspólny język może zatem cywilizować konflikt, ale może go także depolityzować poprzez semantyczne wygładzenie jego realnych podstaw. Właśnie tę linię krytyki rozwijają Oleart i Palomo, ostrzegając przed technosolucjonizmem w demokracji deliberatywnej. Według nich przedstawianie AI jako rozwiązania „problemów demokracji" może odciągać uwagę od masowej polityki, stosunków władzy i ekonomii politycznej samych technologii na rzecz technicznego usprawniania małych procesów konsultacyjnych.

To argument niewygodny, lecz konieczny. Platforma deliberacyjna nie istnieje w próżni, lecz w ramach kapitalizmu cyfrowego. Ma swojego właściciela, infrastrukturę chmurową, model finansowania, dostawców i zależności technologiczne. Pytanie „czy algorytm jest bezstronny?" pozostaje niepełne, jeśli nie zapytamy jednocześnie: kto posiada algorytm i kto może go zmienić? Ryzyko asymetrii władzy zwiększa jeszcze jeden fakt.

Od neutralności algorytmicznej do konstytucjonalizmu deliberacji

Najnowsze badania wskazują, że modele generatywne mogą być skutecznymi narzędziami politycznej perswazji. Eksperymenty dotyczące wyborów prezydenckich w USA w 2024 roku oraz wyborów federalnych w Kanadzie i prezydenckich w Polsce w 2025 roku wykazały istotny wpływ interakcji z systemami AI, opowiadającymi się za konkretnymi kandydatami, na preferencje respondentów. Efekty te okazały się silniejsze niż te typowo obserwowane w przypadku tradycyjnych reklam wideo.

Ustalenie to zmienia etyczny ciężar całej debaty. Technologia, która potrafi pomóc człowiekowi dostrzec racje przeciwnika, może zostać równie dobrze zaprogramowana tak, by przekonywać go do określonej opcji politycznej. Granica między moderacją a perswazją okazuje się zatem technicznie cienka. Jeśli system wybiera kontrargument, który zobaczy uczestnik, już dokonuje selekcji. Jeśli przeredagowuje wypowiedź, wpływa na jej siłę retoryczną. Jeśli streszcza stanowiska grupy, decyduje o tym, które elementy uzna za reprezentatywne. Proponując „wspólny grunt", może podwyższać społeczną cenę dalszego sprzeciwu. AI nie musi wprost nakazywać głosowania na konkretnego kandydata, by skutecznie wpływać na kierunek deliberacji.

Z tego względu pojęcie neutralnego algorytmicznego moderatora należy zastąpić koncepcją moderatora audytowalnego, ograniczonego proceduralnie. Absolutna neutralność jest prawdopodobnie nieosiągalna zarówno dla człowieka, jak i systemu informatycznego. Można jednak projektować kontrolowane warunki bezstronności: publiczne reguły interwencji, testy

symetrii wobec przeciwstawnych poglądów, rejestrowanie wszystkich działań moderatora, niezależne audyty, możliwość odwołania, wersjonowanie modeli, kontrolę driftu, jawność źródeł oraz reprodukowalność eksperymentów. Z perspektywy cyberbezpieczeństwa pojawia się kolejna warstwa: system deliberacyjny staje się infrastrukturą krytyczną procesu demokratycznego w momencie, gdy jego wynik wpływa na referendum, ustawę lub kampanię.

W takim scenariuszu należy brać pod uwagę ataki Sybil, boty, podszywanie się pod uczestników, manipulację tożsamością, prompt injection, zatrucie danych czy ataki na modele klasyfikujące mowę nienawiści, a także wycieki prywatnych transkrypcji i próby strategicznego sterowania podsumowaniami. Pytanie nie brzmi zatem tylko „czy AI jest inteligentna?", lecz „czy cały system jest odporny na przeciwnika zmotywowanego do zmiany rezultatu?".

Równie istotna jest kwestia prywatności. Deliberacja polityczna generuje dane szczególnie wrażliwe: poglądy, emocje, afiliacje, relacje społeczne, wątpliwości, podatność na argumentację oraz historię ewolucji stanowiska. Automatyczna transkrypcja, będąca metodologicznym skarbem dla badacza, staje się jednocześnie potencjalnym archiwum psychopolitycznego profilowania obywateli. W demokracji liberalnej powinna obowiązywać zasada minimalizacji: rejestrowanie wyłącznie tego, co jest niezbędne proceduralnie; ścisłe oddzielenie danych identyfikacyjnych od analitycznych; ograniczenie retencji; kontrola wtórnego wykorzystania danych oraz uniemożliwienie przekształcenia obywatelskiego namysłu w korpus treningowy służący przyszłej komercyjnej perswazji. Inny problem dotyczy języka. AI może radykalnie zwiększyć inkluzywność poprzez tłumaczenie w czasie rzeczywistym, transkrypcję, upraszczanie tekstów eksperckich i adaptację interfejsu dla osób z niepełnosprawnościami. Może tym samym redukować jeden z głównych kosztów Habermasowskiej sfery publicznej: przewagę tych, którzy najlepiej władają dominującym kodem kulturowym. Tłumaczenie nie jest jednak mechanicznym przenoszeniem znaczeń. W polityce różnica między „solidarity" a „welfare", „security" a „order" czy „rights" a „entitlements" może całkowicie zmienić ramę argumentu.

W systemie wielojęzycznym model przestaje być zwykłym tłumaczem, a staje się translatorem normatywnym. Pozwala to wyprowadzić model, który nazwiemy konstytucjonalizmem algorytmicznej deliberacji. Jego podstawowym założeniem jest teza, że im większa jest funkcjonalna rola AI w procesie podejmowania decyzji politycznej, tym bardziej jej reguły powinny przypominać prawo publiczne, a nie regulamin prywatnej platformy. System ograniczony do transkrypcji może podlegać standardom narzędzia pomocniczego. Jednak system dobierający argumenty, decydujący o kolejności mówców i syntetyzujący stanowiska wymaga znacznie silniejszej kontroli. Z kolei rozwiązanie generujące rekomendację, która automatycznie wpływa na decyzję publiczną, zbliża się funkcjonalnie do organu publicznego, co wymaga odpowiednio

wyższego poziomu legitymizacji i odwołalności. Można tu zaproponować zasadę proporcjonalności algorytmicznej władzy: intensywność nadzoru demokratycznego powinna rosnąć wraz z zakresem, w jakim system może modyfikować informację, agendę, uczestnictwo i wynik. Nie każda automatyzacja jest problemem konstytucyjnym – automatyczne odmierzanie 45 sekund nie ma tej samej wagi co automatyczne rozstrzygnięcie o tym, które stanowisko grupy trafi do parlamentu. Projektowanie systemu wymaga więc dekompozycji funkcji, a nie nadawania ogólnej etykiety „AI".

Na tle tych kontrowersji należy przyjrzeć się najbardziej obiecującym wynikom empirycznym. Badanie opublikowane w „Science" z udziałem ponad pięciu tysięcy osób wykazało, że systemy LLM potrafią tworzyć teksty wspólnego stanowiska oceniane przez ludzi wyżej niż te przygotowane przez ludzkich mediatorów. Z kolei eksperyment Stanforda dotyczący automatycznej moderacji pokazał, że szeroka deliberacja online może utrzymać parametry procesu, wywoływać zmiany opinii i prowadzić do późniejszego „deliberative voting". To nie są trywialne rezultaty, jednak nie usprawiedliwiają one przejścia od tezy, że „AI potrafi pomóc w organizacji namysłu", do uznania, iż „AI potrafi demokratycznie rozstrzygać, czym jest dobry namysł". Najnowsza literatura coraz ostrożniej formułuje ten problem.

AI jako narzędzie łagodzenia trylematu deliberatywnego

Autorzy opracowania o AI mediation z 2025-2026 roku przedstawiają sztuczną inteligencję jako potencjalne narzędzie łagodzenia Fishkinowskiego trylematu między szerokim uczestnictwem, znaczącą deliberacją i polityczną równością, ale podkreślają nadal nierozwiązane problemy fairness, transparency, inclusion i participation.

Architektura agregacji zamiast algorytmicznego konsensusu

Nie jest to zatem gotowa technologia ustrojowa, lecz rodzący się obszar eksperymentalnej nauki o instytucjach. Co najbardziej interesujące, AI może zmienić samą naturę problemu skali. Dawna teoria demokracji zakładała niemal oczywisty kompromis: im szerszy krąg uczestników, tym płytsza musi być interakcja. Referendum skaluje się doskonale, lecz nie oferuje deliberacyjnej intensywności; mała rada pozwala na głęboką rozmowę, ale wyklucza większość społeczeństwa. Technologie automatycznej moderacji próbują rozerwać tę zależność poprzez równoległość: milion ludzi nie dyskutuje w jednej gigantycznej sali, lecz uczestniczy w stu tysiącach małych grup, funkcjonujących według wspólnego protokołu. Z punktu widzenia informatyki jest to problem

distributed deliberation. Z perspektywy teorii demokracji pojawia się jednak natychmiast pytanie o agregację wyników tych grup. Czy liczymy średnią? Czy agregujemy rankingi? Czy model syntetyzuje argumenty? Czy mniejszość, powtarzająca się w pięciu procentach grup, ma zostać zachowana w raporcie ogólnym i czy powinno istnieć prawo do publikacji raportu mniejszościowego?

Jak wykrywać rozbieżności regionalne? Jak odróżnić konsensus wynikający z wiedzy od tego, który wyrasta ze wspólnego błędnego założenia? Im skuteczniej technologia rozwiązuje problem logistyczny, tym wyraźniej ujawnia problem teorii wyboru społecznego. Skala nie usuwa polityki – ona przenosi ją do architektury agregacji.

Przyszły system masowej deliberacji nie powinien opierać się na jednym algorytmicznym centrum poznawczym. Bardziej odporny model przypominałby system naukowy: konkurencyjne modele, jawne dane, możliwość replikacji, zespoły czerwone, niezależny audyt, publikacja niezgodności i instytucjonalne prawo do sprzeciwu. W demokracji nie chodzi bowiem o wyprodukowanie jednego „najlepszego podsumowania społeczeństwa", lecz o stworzenie systemu, który pozwoli nam rozumieć zarówno obszary zgody, jak i strukturę własnej niezgody.

Można sformułować tezę znacznie ostrożniejszą od technoentuzjazmu, a zarazem mocniejszą od prostego lęku przed AI: sztuczna inteligencja może być dla demokracji tym, czym druk, statystyka i telekomunikacja były dla wcześniejszych epok – technologią radykalnie rozszerzającą możliwą skalę organizacji wiedzy. Ale tak jak druk stworzył jednocześnie republikę argumentów i masową propagandę, tak AI może jednocześnie skalować deliberację i manipulację. Pytanie nie brzmi: „AI czy demokracja?", lecz: jaka architektura AI wzmacnia autonomię obywatela, a jaka ją absorbuje? Jeżeli system pomaga człowiekowi usłyszeć innych, kontrolować własne uprzedzenia, rozumieć dowody i zachować ostateczną autonomię decyzji, może pełnić rolę infrastruktury deliberacyjnej. Jeżeli natomiast niewidocznie ustala problem, selekcjonuje racje, optymalizuje emocjonalną skuteczność i syntetyzuje wynik, którego obywatel nie potrafi zakwestionować, otrzymujemy nie „mądrego moderatora", lecz algorytmicznego regenta. W tej różnicy mieści się cała przyszłość cyfrowej demokracji deliberatywnej.

Technologia może obniżyć koszt rozmowy, ale nie może zdjąć z człowieka odpowiedzialności za jej reguły. Może pomóc odnajdywać wspólne racje, lecz nie może uczynić konsensusu wartością nadrzędną wobec prawdy, sprawiedliwości i prawa do sprzeciwu. Może wspierać reprezentatywność, ale nie może sama zdefiniować, kogo uznamy za reprezentowanego. Może wykonywać pracę poznawczą, lecz mandat polityczny nadal należy do ludzi.

Dochodzimy tym samym do problemu głębszego niż technologia. Nawet najlepsza procedura deliberacyjna musi odpowiedzieć na pytanie: czyje interesy powinny zostać uwzględnione, zwłaszcza gdy zainteresowani nie mogą uczestniczyć? Dziecko nie zagłosuje nad długiem publicznym, który będzie spłacać. Człowiek przyszłego stulecia nie usiądzie w minipublicznym dotyczącym klimatu. Gatunek zagrożony wyginięciem nie przedstawi własnego argumentu. Mniejszość może zostać liczebnie zdominowana nawet przez bardzo dobrze poinformowaną większość.

Warto zatem przenieść deliberację z obszaru epistemologii w obszar teorii sprawiedliwości. Tam Fishkin spotka Rawlsa, a problem reprezentatywnej próbki zamieni się w pytanie jeszcze trudniejsze: jak zbudować instytucję, która potrafi reprezentować interesy tych, których przy stole nie ma? Deliberacja jako technologia sprawiedliwości: Rawls, mniejszości i polityczna obecność przyszłych pokoleń.

Problem demokracji zaczyna się tam, gdzie kończy się symetria pomiędzy głosującymi a tymi, którzy poniosą skutki ich decyzji. W prostym modelu demokratycznym ci sami ludzie wybierają, ponoszą koszty decyzji, obserwują rezultaty i podczas kolejnych wyborów mogą ukarać sprawców błędu. W rzeczywistym państwie relacja ta jest znacznie bardziej asymetryczna. Dzisiejsza większość może zadłużyć przyszłych podatników, wyczerpać zasoby naturalne, zmienić klimat, wybudować infrastrukturę o stuletnim okresie użytkowania, stworzyć odpady radioaktywne wymagające zabezpieczenia przez okres przekraczający życie wielu pokoleń albo skonstruować system emerytalny, którego pełne konsekwencje ujawnią się długo po zakończeniu kadencji jego twórców.

Bezstronność przez filtrację informacji a nie altruizm

Jednocześnie może podejmować decyzje dotyczące mniejszości, których liczebność nigdy nie pozwoli im przekształcić się w większość wyborczą. W takim świecie zasada "każdy posiada jeden głos" pozostaje kluczowa, ale nie wystarcza do rozwiązania problemu sprawiedliwości.

Autor słusznie kieruje w tym miejscu uwagę ku Rawlsowi, łącząc deliberację z pozycją pierwotną, zasadą maximin, sekwencją czterostopniową oraz problemem przyszłych pokoleń. Trzeba jednak od razu wprowadzić istotne rozróżnienie.

Sondaż Deliberatywny nie jest empiryczną realizacją Rawlsowskiej pozycji pierwotnej, a Rawls nie projektował minipublicznego Fishkina. Związek ten ma charakter analitycznej analogii. Rawls tworzy hipotetyczne urządzenie reprezentacji, które służy identyfikowaniu bezstronnych zasad

sprawiedliwości; Fishkin natomiast buduje empiryczną procedurę, w której realni ludzie zachowują własne tożsamości, wiedzą, kim są i deliberują nad konkretnymi alternatywami. Obie konstrukcje spotykają się w pytaniu o to, jak ograniczyć wpływ przypadkowej pozycji własnej na ocenę reguł, lecz odpowiadają na nie zupełnie innymi metodami.

Rawlsowska pozycja pierwotna jest eksperymentem myślowym. Strony znajdujące się za "zasłoną niewiedzy" nie znają swojego miejsca w społeczeństwie, pozycji klasowej, majątku, płci, rasy, zdolności, światopoglądu ani konkretnej koncepcji dobrego życia. Znają natomiast ogólne fakty dotyczące społeczeństwa, psychologii czy ekonomii – wszystko to, co jest niezbędne do rozumowania o instytucjach.

Konstrukcja ta ma uniemożliwić dopasowanie zasad sprawiedliwości do własnej, przypadkowej przewagi. Nie chodzi zatem o niewiedzę jako wartość samą w sobie, lecz o asymetryczną filtrację informacji: strony mają wiedzieć dużo o świecie, ale nie wiedzieć, które miejsce w tym świecie przypadnie im samym. To przeciwieństwo zwykłej polityki.

Wyborca wie, ile zarabia i jaki podatek zapłaci; wie, gdzie mieszka, czy posiada nieruchomość, z jakich usług publicznych korzysta i jakie rozwiązanie prawdopodobnie obciąży właśnie jego. Rawls pyta: jakie zasady wybrałbyś, gdybyś wiedział, że po usunięciu zasłony możesz znaleźć się po dowolnej stronie instytucji, którą właśnie projektujesz? Projektując system ochrony zdrowia, możesz okazać się człowiekiem zdrowym albo ciężko chorym; tworząc system podatkowy – zamożnym albo ubogim; ustalając prawa mniejszości – członkiem większości albo grupy permanentnie mniejszościowej. Rawls dokonuje czegoś radykalnego: zmienia problem sprawiedliwości z pytania "co jest korzystne dla mnie?" na pytanie "jaką strukturę zaakceptowałabym, nie wiedząc, którą pozycję w niej zajmę?". Nie jest to apel do altruizmu, lecz próbą skonstruowania bezstronności przez usunięcie strategicznej informacji o własnej pozycji.

Źródło określa deliberację jako operacjonalizację Rawlsowskiego maximin. Wymaga to korekty. Rawls rzeczywiście posługuje się argumentacją zbliżoną do reguły maksymalizowania położenia najgorzej sytuowanych, szczególnie przy uzasadnianiu zasady różnicy, ale jego teoria nie sprowadza się do uniwersalnego polecenia: "zawsze wybieraj wariant maksymalizujący minimum". Po pierwsze, podstawowe wolności mają pierwszeństwo leksykalne. Po drugie, równość szans poprzedza zasadę różnicy.

Ekspozycja na różnicę zamiast abstrakcyjnej bezstronności

Po trzecie, maximin pojawia się w specyficznie skonstruowanym problemie wyboru pod zasłoną niewiedzy, a nie jako uniwersalna teoria wszelkich decyzji publicznych. Dla procesu deliberacji ważniejsza od samej nazwy strategii maximin jest zatem zmiana perspektywy i przyjęcie punktu widzenia najmniej uprzywilejowanego uczestnika systemu. Minipubliczne mogą w pewnym stopniu przeprowadzić podobną operację – nie poprzez usunięcie wiedzy o własnej tożsamości, lecz konfrontując uczestników z rzeczywistymi osobami zajmującymi inne miejsca w strukturze społecznej. Rawls osiąga bezstronność poprzez abstrakcję; deliberacja próbuje dochodzić do niej poprzez ekspozycję na różnicę. To dwa odmienne modele epistemologiczne.

Pierwszy można określić mianem modelu depersonalizacji: aby oceniać sprawiedliwie, należy odsunąć informację o tym, kim się jest. Drugi to model poszerzenia perspektywy: zachowaj świadomość własnej tożsamości, lecz dowiedz się, kim są inni i jakie konsekwencje Twoja decyzja wywoła w ich życiu. W społeczeństwach pluralistycznych ten drugi mechanizm może być praktycznie istotniejszy, gdyż rzeczywiste decyzje nie zapadają za zasłoną niewiedzy. Człowiek nie przestaje wiedzieć, że jest przedsiębiorcą, pracownikiem, emerytem czy mieszkańcem konkretnego regionu; może jednak zostać zmuszony do skonfrontowania własnego interesu z interesami innych.

W tym miejscu pojawia się alternatywa utylitarystyczna. Utylitaryzm również dąży do bezstronności, lecz czyni to w inny sposób: nakazuje równo uwzględnić użyteczność wszystkich i wybrać rozwiązanie maksymalizujące sumę lub średnią dobrobytu. Rawls krytykuje ten model m.in. dlatego, że agregacja może teoretycznie dopuścić ogromne korzyści większości kosztem dotkliwych strat mniejszości.

Jego "grubsza" zasłona niewiedzy ma umożliwić ocenę samych preferencji oraz instytucji, które je ukształtowały, zamiast prostego sumowania istniejących pragnień. Spór ten można przełożyć bezpośrednio na projekt deliberacyjny: czy panel obywatelski powinien szukać rozwiązania maksymalizującego sumę korzyści? Czy ma chronić najgorzej położonych lub minimalizować największą krzywdę? A może zapewnić każdemu minimalny próg możliwości życiowych?

Pytania te prowadzą poza Rawlsa, ku prioritaryzmowi, sufficientarianizmowi oraz teorii zdolności Amartyi Sena i Marthy Nussbaum. Nie są to jedynie akademickie wariacje – różne funkcje dobrobytu prowadzą do odmiennych decyzji. Wyobraźmy sobie transformację energetyczną.

Model utylitarystyczny zapyta o całkowity bilans kosztów i korzyści społecznych. Model Rawlsowski silniej skupi się na położeniu tych, którzy ponoszą największy ciężar, oraz na strukturze podstawowych instytucji. Prioritaryzm nada dodatkową wagę poprawie losu osób gorzej sytuowanych, bez konieczności ścisłego maksymalizowania minimum. Sufficientarianizm zapyta, czy wszyscy pozostają powyżej moralnie istotnego progu dobrostanu. Z kolei capability approach będzie badać nie tylko dochód, lecz realne możliwości życia: przemieszczania się, zachowania zdrowia, uczestnictwa i kontroli nad własnym środowiskiem.

Deliberacja jako narzędzie ujawniania alternatyw, a nie gwarant moralności

Deliberacja sama w sobie nie wybiera jednej z tych teorii; jest procedurą, a nie kompletną teorią sprawiedliwości. To rozróżnienie jest kluczowe, gdyż łatwo przypisać dobrze zaprojektowanemu dialogowi magiczną zdolność rozstrzygania sporów normatywnych. Jeśli uczestnicy różnią się w ocenie tego, czy sprawiedliwość wymaga maksymalizacji dobrobytu, ochrony praw, priorytetu dla najmniej uprzywilejowanych czy równości zdolności, dodatkowe informacje mogą doprecyzować naturę konfliktu, ale niekoniecznie go rozwiązać.

Demokracja deliberatywna potrzebuje zatem liberalnych granic. Większość może być doskonale poinformowana, a mimo to podjąć decyzję niesprawiedliwą – wiedza bowiem nie jest tożsama z moralnością. Grupa może znać konsekwencje dyskryminacyjnego rozwiązania i wciąż je preferować. Można tu zaproponować interesujące kryterium tyranii: poważna deprywacja praw człowieka staje się szczególnie trudna do usprawiedliwienia, gdy istnieje osiągalna alternatywa pozwalająca realizować cel publiczny przy mniejszym naruszeniu tych praw. Formułę tę należy traktować jako normatywną rekonstrukcję problemu Fishkina, a nie powszechnie obowiązującą definicję prawną. Jej konstrukcja przypomina jednak logikę konieczności i proporcjonalności: jeśli cel można osiągnąć rozwiązaniem mniej dolegliwym, głębsza ingerencja wymaga wyjątkowo silnego uzasadnienia. Deliberacja może być tutaj technologią ujawniania alternatyw.

Tyrania większości rzadko zaczyna się od świadomej deklaracji chęci skrzywdzenia mniejszości. Częściej wynika z niewiedzy o kosztach, niedostrzegania rozwiązań pośrednich lub traktowania kategorii społecznych w sposób abstrakcyjny. Bezpośrednie spotkanie z przedstawicielem mniejszości może zmienić strukturę informacji – nie dlatego, że osobista historia automatycznie rozstrzyga spór normatywny, lecz ponieważ ujawnia skutki, których większość wcześniej nie uwzględniała.

Przykładem tej możliwości jest bułgarski Deliberative Poll dotyczący sytuacji Romów. Według Fishkina poparcie dla desegregacji szkół romskich wzrosło z 42 do 66 procent, co interpretuje się jako dowód na deliberacyjne uwzględnienie grup marginalizowanych. Wykazaliśmy jednak, że równocześnie wzrosły inne negatywne oceny problemów społecznych związanych z romskimi dzielnicami. Wniosek jest zatem subtelniejszy: deliberacja nie musi produkować idealistycznej tolerancji, ale może oddzielić diagnozę realnego problemu od wyboru polityki segregacyjnej. Jest to poznawczo i normatywnie bardziej interesujące niż prosta narracja o „zmniejszeniu uprzedzeń".

W teorii demokratycznej pojawia się jednak trudniejsze pytanie: kogo w ogóle należy włączyć do demosu? Zasada „all affected interests" głosi w swojej intuicyjnej wersji, że osoby dotknięte decyzją powinny mieć możliwość wpływania na jej kształt. Współczesna literatura wskazuje jednak, że ta pozornie prosta formuła rodzi spory o skalę organizacyjną, rodzaj wpływu, stopień oddziaływania decyzji oraz intensywność wymaganego udziału. Problem ten staje się szczególnie widoczny w kwestiach transgranicznych.

Elektrownia, rzeka, emisje przemysłowe, sztuczna inteligencja, platforma cyfrowa czy polityka podatkowa mogą oddziaływać na ludzi nieposiadających prawa głosu w jurysdykcji podejmującej decyzję. Jeśli kryterium członkostwa w demosie pozostanie wyłącznie obywatelstwem, część osób realnie ponoszących skutki zostanie poza procedurą. Jeśli jednak dopuścimy każdego „dotkniętego", granice demosu stają się zmienne i zależne od konkretnego problemu. Demokracja terytorialna zderza się z demokracją interesariuszy.

Deliberacja jako narzędzie sprawiedliwości międzypokoleniowej

Jeszcze trudniejszy jest przypadek przyszłych pokoleń. Nie mogą one protestować przeciwko naszym decyzjom, nie finansują kampanii, nie posiadają mediów ani reprezentantów wybranych przez siebie i nie są w stanie ukarać dzisiejszego rządu w kolejnych wyborach. Są więc politycznie doskonałą grupą do eksploatacji: ponoszą koszty, nie mając możliwości nałożenia kosztu politycznego na decydenta. Ekonomia polityczna opisuje to zjawisko jako problem krótkiego horyzontu. Politycy maksymalizują prawdopodobieństwo reelekcji w czasie mierzonym kadencjami, podczas gdy skutki wielu polityk publicznych rozciągają się na dekady. Rozdźwięk między polityczną stopą dyskontową a społecznym horyzontem skutków może prowadzić do systematycznego niedoważania przyszłości.

Nie trzeba tu zakładać cynizmu. Nawet odpowiedzialny rząd podlega presji obywateli, którzy odczuwają dzisiejszy ciężar podatku znacznie intensywniej niż abstrakcyjną korzyść, która pojawi się w roku 2050. Klimat jest przypadkiem modelowym, lecz nie jedynym. Dług publiczny, infrastruktura wodna, zasoby glebowe, bioróżnorodność, antybiotykooporność, bezpieczeństwo systemów AI, gospodarka odpadami radioaktywnymi czy finansowanie emerytur posiadają wspólną strukturę temporalną: część korzyści politycznych jest natychmiastowa, koszty zaś odroczone, a osoby czerpiące zyski nie muszą być tymi samymi, które później za nie zapłacą. Rawls próbował włączyć ten problem do teorii sprawiedliwości poprzez just savings principle.

W jego późniejszym ujęciu strony nie wiedzą, do którego pokolenia należą i mają wybrać taką zasadę oszczędzania międzypokoleniowego, jakiej same chciałyby od pokoleń wcześniejszych i jaką byłyby gotowe zastosować wobec następnych. Zasada ta nie wymaga bezgranicznej akumulacji; jej celem jest stworzenie i utrzymanie warunków pozwalających zachować sprawiedliwą strukturę podstawową w czasie. Jest to kwestia kluczowa: sprawiedliwość wobec przyszłości nie oznacza maksymalizacji majątku przyszłych ludzi kosztem dzisiejszych. Gdyby tak było, każde pokolenie mogłoby zostać moralnie zobowiązane do gigantycznego poświęcenia na rzecz jeszcze bogatszej przyszłości. Problem Rawlsowski ma raczej charakter instytucjonalny: co musimy przekazać następcom, aby mogli żyć w ramach sprawiedliwych instytucji i zachować możliwość ich reprodukcji? Współczesna literatura nad sprawiedliwością międzypokoleniową interpretuje to jako przejście od fazy akumulacji do stanu stabilnego („steady state"), w którym sprawiedliwa struktura została dostatecznie zabezpieczona.

Z teorią przyszłych pokoleń wiąże się również problem Parfita, określany jako non-identity problem. Niektóre dzisiejsze decyzje wpływają nie tylko na jakość życia przyszłych osób, ale na to, które osoby w ogóle zaistnieją. Zmiana polityki energetycznej, demograficznej czy gospodarczej może przekształcić cały ciąg relacji, migracji i narodzin. Powstaje paradoks: przyszła osoba nie może stwierdzić, że dana polityka uczyniła jej życie gorszym niż byłoby w innym scenariuszu, ponieważ przy alternatywnej polityce prawdopodobnie w ogóle by się nie urodziła. Z tego powodu sprawiedliwość międzypokoleniowa nie daje się sprowadzić do prostego porównania dobrostanu tych samych ludzi w dwóch różnych wariantach. Wymaga ona myślenia o jakości instytucji, warunkach życia i bezosobowych standardach dobra przyszłych populacji.

Tutaj deliberacja może posiadać przewagę nad zwykłym sondażem. Gdy uczestnik zostaje zmuszony do porównania scenariuszy rozciągniętych na dziesięciolecia, problem przestaje być prostym pytaniem o chęć zapłacenia więcej. Staje się pytaniem: „jaki rozkład kosztów między dzisiejszymi a przyszłymi ludźmi uznajesz za sprawiedliwy, jeśli poznasz długookresowe konsekwencje alternatyw?". Nie usuwa to preferencji czasowej, lecz zmienia jej przedmiot. OECD

właśnie dlatego wskazuje problemy długookresowe i wymagające trudnych kompromisów jako obszary szczególnie nadające się do reprezentatywnych procesów deliberacyjnych.

W danych obejmujących lata 1979–2023 odnotowano 716 takich procesów w 28 państwach OECD, angażujących ponad 80 tysięcy losowo wybranych obywateli; spośród 148 nowych przypadków z lat 2021–2023 aż 60 (czyli 41 procent) dotyczyło środowiska. OECD interpretuje deliberację jako narzędzie szczególnie przydatne wtedy, gdy politycy muszą równoważyć bieżące koszty z przyszłymi korzyściami. Nie oznacza to jednak, że panele obywatelskie są naturalnie „bardziej ekologiczne" niż ogół wyborców – takie twierdzenie byłoby ideologicznym nadużyciem. Deliberujący obywatele mogą odrzucić ambitną politykę klimatyczną, jeśli uznają jej koszty za nadmierne lub technologie za niewiarygodne. Istotą procesu jest coś innego: trudniej utrzymać stanowisko oparte wyłącznie na pomijaniu odroczonych kosztów, gdy materiały i eksperci zmuszają uczestnika do ich jawnego uwzględnienia.

Przykładem może być projekt America in One Room: Climate and Energy oraz koreańskie doświadczenia w zakresie odpadów jądrowych. W pierwszym przypadku istnieją mocne potwierdzenia, że deliberacja zmieniała postawy dotyczące polityki klimatycznej. Interesujący jest również przypadek Shin-Gori z Korei Południowej, gdzie panel poparł wznowienie budowy reaktorów (59,5 procent), a jednocześnie opowiedział się za długookresowym ograniczaniem zależności kraju od energetyki jądrowej (53,2 procent). To rozdzielenie krótkiego i długiego horyzontu jest szczególnie istotne.

Koreański panel potrafił bowiem sformułować spójny wniosek: dokończmy kosztowną inwestycję, która jest już zaawansowana, ale w przyszłości zmniejszajmy zależność od tej technologii. Taki wynik nie mieści się łatwo w binarnej strukturze partyjnego hasła „za atomem" albo „przeciw atomowi".

Pluralizm instytucjonalny jako gwarancja widzialności przyszłych pokoleń

Deliberacja pozwala ujawnić czasową wielowymiarowość preferencji: człowiek może popierać decyzję A dzisiaj, a jednocześnie strategię B dla systemu w długim okresie, zwłaszcza gdy koszty przejścia są asymetryczne. Współczesne państwa zaczynają odpowiadać na problem przyszłych pokoleń również poza ramami minipublicznych. Przykładem jest Walia, która na mocy Well-being of Future Generations Act powołała Future Generations Commissioner – instytucję odpowiedzialną za reprezentowanie interesów przyszłych generacji, doradzanie organom publicznym oraz ocenę,

czy państwo uwzględnia długofalowe skutki swoich decyzji. Rozwiązanie to opiera się na innej filozofii niż ta proponowana przez Fishkina.

Zamiast powoływać reprezentatywny mikrokosmos społeczeństwa, tworzy ono instytucjonalnego rzecznika nieobecnych. W tym kontekście można wyróżnić kilka konkurencyjnych modeli reprezentowania przyszłości: model konstytucyjny, który wpisuje obowiązek długookresowości w normy najwyższej rangi; model rzeczniczy, powołujący urząd chroniący przyszłe interesy; model ekspercki, oparty na radach naukowych i fiskalnych oceniających skutki w czasie; model deliberacyjny, angażujący obywateli w proces ważenia kosztów obecnych i przyszłych; model parlamentarny, ustanawiający komisje ds. przyszłości, oraz model sądowy, pozwalający traktować zobowiązania środowiskowe jako normy podlegające kontroli prawnej.

Żaden z tych modeli nie jest samodzielnie kompletny. Rzecznik przyszłych pokoleń dysponuje silnym mandatem normatywnym, lecz słabą reprezentacją demokratyczną. Panel obywatelski zapewnia społeczną różnorodność, ale jego uczestnicy wciąż należą wyłącznie do pokolenia obecnego. Ekspert potrafi modelować skutki do 2100 roku, jednak samo modelowanie nie rozstrzyga, jakie ryzyko moralnie wolno przenieść na ludzi urodzonych za pół wieku. Sąd może chronić prawa, lecz nie powinien samodzielnie prowadzić polityki klimatycznej, a parlament posiada mandat, którego stopa dyskontowa jest skracana przez cykl wyborczy.

Najbardziej odporny system wymaga prawdopodobnie instytucjonalnego pluralizmu reprezentacji przyszłości, co stanowi analogię do teorii redundancji demokratycznej. Przyszłość nie może zależeć od jednego strażnika. Optymalne rozwiązanie zakładałoby splot wielu mechanizmów: obowiązkową ocenę skutków międzypokoleniowych dla ustaw, niezależny urząd rzecznika, panele deliberacyjne przy decyzjach o bardzo długim horyzoncie oraz publikację scenariuszy eksperckich wymuszających na parlamencie uzasadnienie przeniesienia kosztów na przyszłe generacje. Nie jest konieczne, by każda z tych instytucji posiadała prawo weta; kluczowe jest natomiast to, aby przyszły koszt przestał być politycznie niewidzialny.

OECD odnotowuje obecnie rosnącą instytucjonalizację deliberacji w obszarach długookresowych. Między 2020 a 2023 rokiem liczba stałych mechanizmów deliberacyjnych wzrosła z 22 do 41, przy czym wiele z nich dotyczy zielonej transformacji – czego przykładem są permanentne zgromadzenia klimatyczne w Brukseli czy Mediolanie. Proces ten jest istotny, gdyż jednorazowy panel generuje jedynie punktową rekomendację, podczas gdy kryzys klimatyczny i transformacja infrastrukturalna ewoluują. Instytucja stała pozwala budować pamięć deliberacyjną przekraczającą jeden cykl polityczny.

Zasada zachowania opcji przyszłości i wielostopniowa architektura namysłu

Przyszłe pokolenia stawiają przed teorią demokracji problem jeszcze bardziej złożony. Nie mogą być reprezentowane statystycznie, ponieważ nie istnieją w populacji, z której losujemy próbę. Nie można ich zatem włączyć do minipublicznego na takich samych zasadach jak młodych, starszych, mieszkańców wsi czy mniejszości etniczne.

Ich obecność musi być konstruowana normatywnie. Możemy posłużyć się scenariuszami typu: „wyobraź sobie, że rodzisz się w 2080 roku", ustanowić w grupie rolę „adwokata przyszłości" lub wymagać odrębnej oceny wpływu decyzji na przyszłe generacje. Możemy prezentować uczestnikom modele długookresowe czy zapraszać młodych ludzi jako najbliższych temporalnie reprezentantów dalszej przyszłości. Wszystkie te rozwiązania pozostają jednak substytutami, a nie realnym przedstawicielstwem. Człowiek żyjący dziś nigdy nie przestaje być beneficjentem lub ofiarą teraźniejszości.

Ta niemożność pełnej reprezentacji powinna skłaniać nie do rezygnacji, lecz do epistemicznej pokory. Nieobecny nie może skorygować naszej interpretacji jego interesu. Właśnie dlatego decyzje nieodwracalne wymagają wyższego standardu ostrożności niż te łatwo odwracalne. Zasada ta zbliża deliberację do teorii opcji realnych i precautionary principle: gdy przyszłe pokolenie nie będzie mogło odzyskać utraconego zasobu, gatunku, stabilności klimatycznej czy bezpieczeństwa wynikającego z decyzji technologicznej, wartość zachowania przyszłej przestrzeni wyboru rośnie. Możemy określić to mianem zasady zachowania opcji przyszłości. Nie oznacza ona konserwowania status quo za wszelką cenę, gdyż brak działania również może bezpowrotnie zamknąć pewne możliwości. Ocieplenie klimatu, degradacja wód, narastający dług inwestycyjny czy niekontrolowany rozwój ryzykownej technologii mogą stworzyć path dependence równie silną jak aktywna decyzja. Pytanie o sprawiedliwość brzmi zatem: które dzisiejsze wybory nieodwracalnie ograniczają zdolność przyszłych ludzi do projektowania własnego życia?

Warto w tym miejscu wrócić do Rawlsowskiej four-stage sequence. Choć notatki sugerują jej związek z deliberacją, Rawls rozumiał ją precyzyjniej jako proces stopniowego przechodzenia od wyboru ogólnych zasad sprawiedliwości, przez etap konstytucyjny i ustawodawczy, aż po zastosowanie reguł do konkretnych przypadków – przy sukcesywnym odsłanianiu informacji o realnym społeczeństwie. Jest to fascynujący model dla teorii deliberacyjnej, gdyż sugeruje, że różne pytania wymagają różnych poziomów abstrakcji i odmiennych zbiorów danych.

Projektując konstytucję, powinniśmy pytać o reguły, których jesteśmy gotowi przestrzegać niezależnie od tego, kto wygra najbliższe wybory. Przy ustawie potrzebujemy szerszej wiedzy o gospodarce i strukturze społecznej, a przy wykonaniu – danych lokalnych i indywidualnych. Jedną z patologii współczesnej polityki jest mieszanie tych poziomów: reguły konstytucyjne projektuje się pod bieżącą większość, a sprawy techniczne zamieniają się w plemienne plebiscyty. Rawlsowska sekwencja przypomina, że sprawiedliwość wymaga odpowiedniego dystansu poznawczego względem rodzaju podejmowanej decyzji.

Prowadzi to do projektu deliberacji wielostopniowej. Najpierw minipubliczne mogłoby deliberować nad wartościami i zasadami ograniczającymi rozwiązanie; następnie eksperci przedstawiliby warianty zgodne z tymi warunkami, a kolejny panel porównałby skutki dystrybucyjne. Parlament podejmowałby ostateczną decyzję, po której – w określonym czasie – następowałaby deliberacyjna ocena skutków. Nie byłaby to jedna rozmowa, lecz łańcuch refleksyjny, w którym normy, dane i doświadczenie są stopniowo integrowane.

Takie podejście pozwala również lepiej zarządzać niepewnością. W debacie publicznej ekspert często zmuszany jest do udawania pewności, ponieważ komunikat „model daje szeroki przedział" przegrywa medialnie z twierdzeniem „wiemy, co się stanie". Deliberacja może uczynić niepewność przedmiotem jawnego wyboru. Uczestnicy powinni otrzymywać nie tylko prognozę centralną, lecz rozkład scenariuszy, poziom ufności i opis tego, które skutki są odwracalne. Sprawiedliwość międzypokoleniowa wymaga bowiem nie tylko analizy oczekiwanej wartości skutku, ale także badania ogona rozkładu ryzyka. Polityka może mieć umiarkowanie dodatnią wartość oczekiwaną, a jednocześnie niewielkie prawdopodobieństwo katastrofalnego i nieodwracalnego rezultatu. Pokolenie obecne może czerpać korzyści, podczas gdy przyszłe ponosi ryzyko skrajne. W takich sytuacjach klasyczna agregacja użyteczności, Rawlsowska ostrożność i zasada przezorności mogą prowadzić do sprzecznych rekomendacji. Deliberacja powinna ujawniać ten konflikt wartości, zamiast ukrywać go pod jedną liczbą.

Dobrze zaprojektowany minipubliczny jest potencjalnie symulatorem pozycji moralnych, a nie po prostu reprezentacją socjodemograficzną.

Deliberacja jako technologia widzialności nieobecnych

Oprócz pytania o to, czy w próbie reprezentatywnym jest odsetek osób młodych i starszych, należy zapytać: czy procedura zmusza uczestników do rozważenia perspektywy najgorzej sytuowanych, przyszłych podatników, mniejszości, mieszkańców terenów dotkniętych inwestycją oraz ludzi

żyjących już po zakończeniu kadencji dzisiejszych polityków? Tu ponownie powraca kontrowersja Lafont.

Jeżeli panel posiada szczególne warunki pozwalające uwzględnić te perspektywy, może dojść do wniosków odmiennych od poglądów większości populacji. Czy rezultat jest przez to bardziej sprawiedliwy? Nie automatycznie. Minipublic nie zyskuje moralnej nieomylności dzięki samemu losowaniu. Jego wartość polega na tym, że może ujawnić społeczeństwu koszt moralny rozwiązania, którego standardowa kampania polityczna nie miała bodźca ukazywać. Ostateczna legitymizacja powinna zatem wynikać z powrotu argumentów do szerszego obiegu publicznego.

Przyszłe instytucje deliberacyjne nie powinny ograniczać się do publikowania suchej „rekomendacji 68 procent za". Byłoby to ponowne sprowadzenie deliberacji do poziomu sondażu. Powinny one publikować mapę racji: jakie argumenty uznano za kluczowe, jakie koszty odkryto, których wartości nie udało się pogodzić, jakie stanowisko zajęła mniejszość panelu, jak ewoluowały poglądy po kolejnych fazach informacyjnych i jakie niepewności pozostały nierozstrzygnięte. Dla społeczeństwa ważniejsza od samego wyniku może być wiedza, dlaczego reprezentatywna grupa doszła do określonego sądu.

Jest to odpowiedź na problem „black box" Lafont. Im większe znaczenie minipublicznego, tym mniej jego rozumowanie może pozostawać czarną skrzynką. Demokratycznie użyteczna deliberacja musi eksportować racje, nie tylko werdykty. W rezultacie możemy sformułować tezę znacznie bardziej rygorystyczną niż ta, że deliberacja po prostu „sprzyja sprawiedliwości". Sprzyja jej tylko wtedy, gdy konstrukcja procesu zwiększa widzialność konsekwencji dla osób słabo reprezentowanych przez standardowe bodźce wyborcze, wymusza porównanie realnych alternatyw, chroni prawa przed czystą agregacją większościową, pozwala analizować długookresowe i niepewne skutki oraz pozostawia rezultat otwarty na szerszą krytykę publiczną. Nie jest więc maszyną produkującą sprawiedliwość, lecz technologią poszerzania pola moralnie relewantnej informacji.

W ten sposób można przekształcić maksymę, według której sprawiedliwość potrzebuje spotkania z innością. Sprawiedliwość potrzebuje także spotkania z nieobecnością: z człowiekiem, który nie ma wystarczającej siły głosu, z dzieckiem, które jeszcze nie głosuje, z przyszłym obywatelem, który jeszcze się nie urodził, oraz z konsekwencją, która nie zdąży ujawnić się przed następnymi wyborami. Demokracja teraźniejszości łatwo staje się ustrojem żyjących przeciwko nieobecnym. Demokracja deliberatywna może stać się dojrzała dopiero wtedy, gdy nauczy się uczynić politycznie widzialnym interes tych, którzy nie mogą wejść do sali.

To prowadzi nas bezpośrednio do Polski. Polski porządek konstytucyjny posiada już wiele elementów, z których można budować podobną architekturę: petycję, referendum, konsultacje,

wysłuchanie publiczne, obywatelską inicjatywę ustawodawczą, samorząd terytorialny, komisje parlamentarne oraz rozwijające się praktyki paneli obywatelskich.

Pytanie nie brzmi więc, czy trzeba importować całkowicie obcy model. Brzmi: gdzie w istniejącym polskim obiegu decyzji znajdują się miejsca, w których głos obywatela jest już dopuszczony, ale nie zapewniono jeszcze warunków do przejścia od reakcji do przemyślanego sądu? Poświęcimy się Polsce jako potencjalnemu laboratorium przejścia od demokracji przedstawicielskiej do reprezentacji kontrolowanej przez obywatelski namysł.

> Demokracja teraźniejszości zbyt często okazuje się ustrojem żyjących przeciwko nieobecnym. Prawdziwa dojrzałość cyfrowego namysłu nadejdzie nie wtedy, gdy algorytmy znajdą wspólny język dla wszystkich, lecz gdy nauczymy się za ich pomocą uczynić politycznie widzialnych tych, którzy nigdy nie wejdą do sali obrad. Czy potrafimy zatem zaprojektować system, w którym głos nieistniejącego jeszcze obywatela będzie ważył tyle, co obecny interes większości?

Inspiracje

[1] 

"Can Deliberation Cure the Ills of Democracy" James S Fishkin

ISBN: 9780198944416

James S. Fishkin (ur. 1948) to wybitny amerykański politolog i profesor znany z pionierskich prac w dziedzinie teorii demokracji i demokracji deliberatywnej. Obecnie pełni funkcję Janet M. Peck Chair in International Communication na Uniwersytecie Stanforda, gdzie kieruje również Centrum Demokracji Deliberatywnej. Fishkin jest najbardziej znany z wynalezienia „sondażu deliberatywnego", metody badawczej mającej na celu badanie świadomej opinii publicznej poprzez gromadzenie reprezentatywnych prób obywateli w celu omówienia kwestii politycznych. Jego praca naukowa koncentruje się na styku teorii polityki i badań empirycznych, a w szczególności bada, w jaki sposób ustrukturyzowana deliberacja może usprawnić procesy demokratyczne i złagodzić polaryzację polityczną. Jest autorem wielu wpływowych książek i artykułów na temat reform demokratycznych, opinii publicznej i projektowania instytucji politycznych, zyskując międzynarodowe uznanie za swoje wysiłki na rzecz uczynienia demokracji bardziej inkluzywną i deliberatywną.

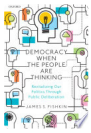
James S. Fishkin (born 1948) is a prominent American political scientist and professor known for his pioneering work in democratic theory and deliberative democracy. He currently serves as the Janet M. Peck Chair in International Communication at Stanford University, where he also directs the Center for Deliberative Democracy. Fishkin is best known for inventing the 'Deliberative Poll,' a research method designed to gauge informed public opinion by bringing together representative samples of citizens to discuss policy issues. His academic work focuses on the intersection of political theory and empirical research, specifically exploring how structured deliberation can improve democratic processes and mitigate political polarization. He has authored numerous influential books and articles on democratic reform, public opinion, and the design of political institutions, earning international recognition for his efforts to make democracy more inclusive and deliberative.

Stanford University

Literatura uzupełniająca

Książki

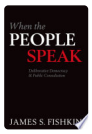
Książki autora i dzieła przywołane w tekście:



[1] "Democracy When the People Are Thinking"

James S. Fishkin

2018 | Oxford University Press | ISBN: 9780192551894



[2] "When the People Speak"

James S. Fishkin

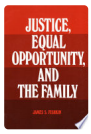
2011 | Oxford University Press | ISBN: 9780199604432



[3] "The Voice of the People"

James S. Fishkin

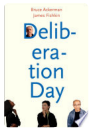
1995 | ISBN: 9780300065565



[4] "Justice, Equal Opportunity, and the Family"

James S. Fishkin

1983 | Yale University Press | ISBN: 9780300032499

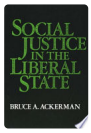


[5] "Deliberation Day"

Bruce Ackerman, James S. Fishkin

2008 | Yale University Press | ISBN: 9780300127027

Książki autorów przywoływanych i literatura pokrewna:



[6] "Social justice in the liberal state."

Bruce Ackerman

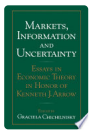
Yale University Press | ISBN: 9780300027570



[7] "The Past as Future"

Jürgen Habermas

U of Nebraska Press | ISBN: 9780803272668



[8] "Markets, Information and Uncertainty"

Kenneth Joseph Arrow, Graciela Chichilnisky

1999 | Cambridge University Press | ISBN: 9780521553551



[9] "Essays in the Theory of Risk-bearing"

Kenneth Joseph Arrow

1971



[10] "For a Left Populism"

Chantal Mouffe

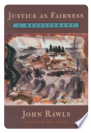
2018 | Verso Books | ISBN: 9781786637550



[11] "The Democratic Paradox"

Chantal Mouffe

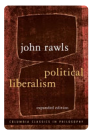
2000 | Verso | ISBN: 9781859847589



[12] "Justice as Fairness"

John Rawls

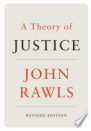
Harvard University Press | ISBN: 9780674244290



[13] "Political Liberalism"

John Rawls

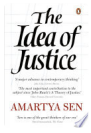
Columbia University Press | ISBN: 9780231130882



[14] "A Theory of Justice"

John Rawls

Harvard University Press | ISBN: 9780674042582



[15] "The Idea of Justice"

Amartya Sen

2010 | Penguin UK | ISBN: 9780141969480

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Energy Policy and the Further Future: The Identity Problem"

Derek Parfit

2010 | Climate Ethics | DOI: 10.1093/oso/9780195399622.003.0015

Źródło

[2] "Review of Emotional communication and therapeutic change: Understanding psychotherapy through multiple code theory."

Tilmann Habermas

2023 | Psychoanalytic Psychology | DOI: 10.1037/pap0000463

Źródło

[3] **"Essays in the Theory of Risk-Bearing."**

Nils H. Hakansson, Kenneth Joseph Arrow

1972 | The Journal of Finance | DOI: 10.2307/2978877

Źródło

[4] **"The European Union's 'phantom menaces': What can the European political system learn from the institutional design and the fall of Star Wars' Old Galactic Republic?"**

Lorenzo Díez-Picazo, Nicolás Palomo Hernández

2025 | DOI: 10.17613/20w11-xc144

Źródło

[5] **"Is the Economist Intelligence Unit's Democracy Index little more than a joke indicator?"**

Nicolás Palomo Hernández

2023 | Knowledge Commons (Lakehead University) | DOI: 10.17613/efrxn-nv821

Źródło

[6] **"Deliberative Democracy or Agonistic Pluralism?"**

Chantal Mouffe

2026 | Models of Deliberative Democracy | DOI: 10.4324/9781003763987-14

Źródło

[7] **"Dimensions of radical democracy : pluralism, citizenship, community"**

Chantal Mouffe

1992

Źródło

[8] **"Democracy as Agonistic Pluralism"**

Chantal Mouffe

2017 | Rewriting Democracy | DOI: 10.4324/9781315244167-3

Źródło

[9] **"Democracy, human rights and cosmopolitanism: an agonistic approach"**

Chantal Mouffe

2014 | The Meanings of Rights | DOI: 10.1017/cbo9781139227124.014

Źródło

[10] "Contemporary Political Philosophy and Religion : between Public Reason and Pluralism"

J. Rawls, R. Audi, S. Moller

2021

Źródło

[11] "Future generations: Further problems"

Derek Parfit

1982 | DigitalGeorgetown (Georgetown University Library)

Źródło

[12] "Future People, the Non-Identity Problem, and Person-Affecting Principles"

Derek Parfit

2017 | Philosophy &amp Public Affairs | DOI: 10.1111/papa.12088

Źródło

Artykuły naukowe autora:

[13] "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion"

James S. Fishkin, Robert C. Luskin

2005 | Acta Politica | DOI: 10.1057/palgrave.ap.5500121

Źródło

[14] "Considered Opinions: Deliberative Polling in Britain"

Robert C. Luskin, James S. Fishkin, Roger M. Jowell

2002 | British Journal of Political Science | DOI: 10.1017/s0007123402000194

Źródło

[15] "Democracy When the People Are Thinking"

James S. Fishkin

2018 | Oxford University Press eBooks | DOI: 10.1093/oso/9780198820291.001.0001

Źródło

[16] "When the People Speak"

James S. Fishkin

2011 | Oxford University Press eBooks | DOI: 10.1093/acprof:osobl/9780199604432.001.0001

Źródło

[17] "Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on “America in One Room”"

James S. Fishkin, Alice Siu, Larry Jay Diamond, Norman M. Bradburn

2021 | American Political Science Review | DOI: 10.1017/s0003055421000642

Źródło

[18] "Democracy and Deliberation: New Directions for Democratic Reform"

Stephen L. Elkin, James S. Fishkin

1993 | Journal of Policy Analysis and Management | DOI: 10.2307/3325243

Źródło

[19] "Disaggregating Deliberation’s Effects: An Experiment within a Deliberative Poll"

Cynthia Farrar, James S. Fishkin, DONALD PHILIP GREEN, Christian List, Robert C. Luskin

2010 | British Journal of Political Science | DOI: 10.1017/s0007123409990433

Źródło

[20] "Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls"

Christian List, Robert C. Luskin, James S. Fishkin, Iain McLean

2012 | The Journal of Politics | DOI: 10.1017/s0022381612000886

Źródło

[21] "Deliberative polling and public consultation"

James S. Fishkin, RC Luskin, Roger M. Jowell

2000 | Parliamentary Affairs | DOI: 10.1093/pa/53.4.657

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: 4c082c85-70cc-45d0-8ddf-42a79831248b