Utopia jako laboratorium moralne w kontekście The Challenge of the Future A. Graylinga

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje koncepcję utopii nie jako nierealnej fantazji, lecz jako laboratorium filozofii politycznej, które pozwala testować instytucjonalne koszty konkretnych hierarchii wartości. Autor wskazuje, że utopia i dystopia są narzędziami służącymi do ujawniania skutków konsekwentnego stosowania zasad organizacji świata. Poprzez odwołania do Morusa i Bacona, artykuł ukazuje, jak wizje idealnego ustroju stają się lustrem dla rzeczywistości, podważając oczywistość obecnych struktur społecznych. Szczególną uwagę poświęcono zagrożeniom płynącym z technokratycznej optymalizacji, gdzie nauka i technologia przestają być narzędziami poznania, a stają się instrumentami kontroli i wytwarzania przyszłości, co bezpośrednio zagraża ludzkiej autonomii w dobie cyfryzacji.

The text analyzes the concept of utopia not as an unrealistic fantasy, but as a laboratory of political philosophy that allows for testing the institutional costs of specific value hierarchies. The author indicates that utopia and dystopia are tools used to reveal the effects of consistently applying principles of world organization. Through references to More and Bacon, the article shows how visions of an ideal system become a mirror for reality, challenging the obviousness of current social structures. Particular attention is paid to the threats arising from technocratic optimization, where science and technology cease to be tools of knowledge and instead become instruments of control and the manufacturing of the future, directly threatening human autonomy in the age of digitalization.

Artykuł analizuje utopię i dystopię nie jako literackie fantazje, lecz jako rygorystyczne laboratoria filozoficzne, które pozwalają badać instytucjonalne koszty konkretnych

wartości. Autor stawia tezę, że współczesna technoutopia, obiecując świat pozbawiony tarcia i zoptymalizowany pod kątem efektywności, niesie ze sobą fundamentalne zagrożenie dla ludzkiej autonomii. Głównym ryzykiem nie jest tu wroga superinteligencja, lecz „pułapka skuteczności" – sytuacja, w której wygoda i precyzyjne algorytmiczne rekomendacje prowadzą do dobrowolnej delegacji sprawstwa i zaniku zdolności do samodzielnego osądu. Tekst ostrzega przed mechanizmem eksternalizacji kosztów (na przykładzie paraboli Omelas), wskazując, że cyfrowy dobrobyt często opiera się na ukrytym cierpieniu lub wykluczeniu. W konkluzji autor postuluje zastąpienie marzenia o nieomylnym systemie ideałem „korygowalności", gdzie wolność definiowana jest jako prawo do bycia nieoptymalnym i możliwość sprzeciwu wobec dyktatu technicznej racjonalności.

SŁOWA KLUCZOWE

utopia jako laboratorium moralne | technokratyczna optymalizacja | przymus infrastrukturalny | asymetria epistemiczna | Mother Goddess AI | transhumanistyczna eutopia | korygowalność systemu | geografia kosztów | homogenizacja społeczna | paternalizm technologiczny | delegowanie podmiotowości | infrastruktura podmiotowości | pluralizm wartości | mechanizmy kontroli

Utopia jako laboratorium kosztów instytucjonalnych

Utopia jako laboratorium moralne: od dobrego miejsca do doskonałej kontroli. Podczas gdy aksjologia odpowiada na pytanie, co jest warte chcenia, utopia wykonuje kolejny krok: próbuje wyobrazić sobie świat, w którym wartości zostały już przełożone na instytucje. Literatura utopijna jest zatem czymś znacznie poważniejszym niż zbiorem fantazji o szczęśliwych wyspach, sprawiedliwych republikach czy technologicznych rajach. Stanowi laboratorium filozofii politycznej, gdyż zmusza autora do przedstawienia konsekwencji konkretnej hierarchii wartości. Jeśli najważniejsza jest równość, trzeba pokazać, jak ograniczyć nierówności; jeśli wolność – jak poradzić sobie z jej konfliktami; jeśli bezpieczeństwo – jakie kompetencje przekazać władzy; jeśli wiedza – kto ma nią zarządzać; a jeśli szczęście – co zrobić z jednostkami odmawiającymi uczestnictwa w zbiorowym projekcie. Utopia jest zatem testem nie tylko marzenia, lecz jego instytucjonalnego kosztu.

Grayling wyraźnie nadaje tej tradycji taką funkcję, traktując literaturę jako „laboratorium wyobraźni", które pozwala testować scenariusze – od rozkwitu po technologiczne niewolnictwo. Historyczna tradycja utopijna prowadzi tu od Morusa i Bacona, przez alternatywne wizje społeczne, aż po XX-wieczne dystopie Huxleya i Orwella. Nie należy jednak czytać tej historii jako prostego przejścia od dawnego optymizmu do współczesnego pesymizmu. Utopia i dystopia są raczej bliźniaczymi narzędziami jednego rodzaju rozumowania: obie wyostrzają pewną zasadę organizacji świata, aby ujawnić skutki jej konsekwentnego zastosowania. Różnią się przede wszystkim tym, czy ostateczny obraz ma wzbudzić pragnienie, czy ostrzeżenie.

Już samo słowo „utopia" zawiera tę dwuznaczność. Tomasz Morus, publikując w 1516 roku dzieło, które nadało nazwę całemu gatunkowi, wykorzystał grę pomiędzy greckim ou-topos – miejscem, którego nie ma – a eu-topos – miejscem dobrym. Grayling czyni z tej etymologicznej ambiwalencji symbol problemu: dobre miejsce może okazać się nie-miejscem właśnie dlatego, że nie istnieje bez kosztów, kompromisów i konfliktów właściwych rzeczywistym społeczeństwom. Utopia nie jest więc tylko projektem idealnego ustroju; jest pytaniem o to, czy „idealny ustrój" jest pojęciem spójnym, skoro społeczeństwo musi godzić dobra częściowo niewspółmierne.

Morusowska Utopia pozostaje pod tym względem dziełem zadziwiająco nowoczesnym. Z jednej strony przedstawia społeczeństwo pozbawione prywatnej akumulacji bogactwa, zorganizowane wokół wspólnej własności, pracy i edukacji; z drugiej zachowuje instytucje i praktyki dalekie od emancypacyjnego ideału. Nie ma zatem jednej odpowiedzi na pytanie, czy Morus opisuje system, który sam chciałby wprowadzić, czy też używa fikcji jako instrumentu ironii i krytyki europejskiego społeczeństwa. Ta niejednoznaczność jest metodologicznie cenna.

Utopia nie musi być programem. Może być lustrem, w którym rzeczywistość staje się dziwna właśnie dlatego, że zostaje zestawiona z systemem opartym na odmiennych założeniach. To odwrócenie perspektywy jest podstawowym mechanizmem myślenia utopijnego. To, co codzienne, wydaje się naturalne tylko dopóki nie zobaczymy świata, w którym ten sam problem rozwiązano inaczej. Prywatna własność, praca najemna, rodzina, pieniądz czy model miasta mogą uchodzić za oczywiste jedynie dlatego, że funkcjonujemy wewnątrz określonej konfiguracji historycznej. Utopia zawiesza domniemanie konieczności.

Pyta: gdybyśmy zaczynali od początku, czy zaprojektowalibyśmy instytucję w ten sam sposób? Samo postawienie tego pytania jest politycznie destabilizujące, ponieważ ujawnia różnicę między „tak jest" a „tak musi być". Nieprzypadkowo utopie powstają najintensywniej w epokach głębokiej transformacji. Gwałtowna zmiana zwiększa świadomość historyczności instytucji. Gdy świat zmienia się powoli, porządek społeczny może sprawiać wrażenie przedłużenia natury. Jednak gdy reformacja, odkrycia geograficzne, rewolucja naukowa czy cyfryzacja w ciągu jednego pokolenia

przebudowują sposób życia, człowiek dostrzega, że społeczeństwo może przyjmować wiele form. Utopia rodzi się zatem nie tylko z nadziei na lepszy świat, lecz z doświadczenia, że świat zastany nie jest jedynym możliwym.

Technokratyczna optymalizacja jako zagrożenie dla autonomii

Francis Bacon przesuwa to rozumowanie w kierunku, który okaże się szczególnie istotny dla XXI wieku. New Atlantis stanowi jedną z klasycznych wizji społeczeństwa, w którym nauka staje się instytucjonalnym fundamentem rozwoju. Bacon jawi się tu jako prekursor technoptymizmu, postrzegający wiedzę jako siłę pozwalającą opanować naturę. W jego projekcie pojawia się jednak coś więcej niż pochwała eksperymentu. Wiedza zostaje zorganizowana i podporządkowana systematycznej produkcji nowych możliwości; nauka przestaje być jedynie kontemplacją prawdy, a staje się technologią wytwarzania przyszłości. To przesunięcie stanowi jedno z głównych źródeł nowoczesności.

Przednowoczesne marzenia o doskonałym społeczeństwie opierały się często na ideałach prawa, obyczaju, religii lub mądrości władcy. Utopia technologiczna wprowadza nową perspektywę: społeczeństwo można ulepszyć poprzez systematyczne rozszerzanie kontroli nad przyrodą. Choroba staje się problemem medycznym, głód agronomicznym, odległość transportowym, a brak informacji komunikacyjnym.

To ogromny cywilizacyjny sukces, którego krytyka nie może uczciwie negować. Wiele form cierpienia ograniczono właśnie dlatego, że przestano traktować je jako przeznaczenie, a zaczęto jako problem techniczny. Tutaj jednak rodzi się charakterystyczna dla nowoczesności pokusa: skoro tak wiele ograniczeń można rozwiązać technicznie, być może zasadnicze problemy społeczne są w istocie problemami inżynieryjnymi. Nieefektywność można zoptymalizować, przestępczość przewidzieć, zachowania skorygować, a konflikty zredukować odpowiednią architekturą bodźców, czyniąc społeczeństwo bardziej racjonalnym dzięki lepszym danym. Technokratyczna utopia nie musi obiecywać pałaców ze złota; wystarczy, że obieca system pozbawiony tarcia.

Tymczasem nie każde społeczne „tarcie” jest usterką. Spór polityczny może być kosztem pluralizmu, proceduralna powolność ceną praworządności, możliwość błędu warunkiem autonomii, a nieprzewidywalność zachowania kosztem wolności. System zoptymalizowany według jednego kryterium może zniszczyć dobra, które ujawniają się dopiero w konflikcie z tym kryterium.

Utopia technologiczna staje się więc szczególnie niebezpieczna, gdy polityczne napięcie interpretuje jako problem obliczeniowy, a człowieka jako zmienną wymagającą właściwego skalibrowania.

Jonathan Swift reprezentuje przeciwstawną funkcję wyobraźni. Satyra nie projektuje świata pożądanego, lecz poprzez deformację ujawnia groteskę istniejących pretensji do rozumu i wielkości. Podróże Gulliwera przypominają, że żaden system nie zostaje automatycznie zabezpieczony przed ludzką próżnością, interesem i konfliktem przez samą elegancję jego konstrukcji. Swift wnosi do dyskusji o przyszłości głęboki sceptycyzm wobec natury ludzkiej: instytucja działa nie w świecie idealnych parametrów, lecz w rzeczywistości zamieszkanej przez istoty zdolne do samousprawiedliwienia, rywalizacji, resentymentu i absurdu.

Ta lekcja może być ważniejsza dla rozwoju sztucznej inteligencji niż najbardziej futurystyczne wizje. Znaczna część dyskursu technologicznego zakłada bowiem milcząco, że jeśli system będzie dostatecznie inteligentny lub optymalny, znikną konflikty wynikające z ludzkiej niedoskonałości. Tymczasem technologia wchodzi w istniejące struktury interesów. System predykcyjny nie usuwa walki o władzę; może ją jedynie wyposażyć w nowe narzędzie. Automatyzacja nie rozwiązuje sporu o dystrybucję, lecz zmienia wielkość i położenie zasobu podlegającego podziałowi.

Generowanie większej ilości informacji nie eliminuje manipulacji, a może zwiększyć jej skalę. Rozum instrumentalny poszerza zatem możliwości zarówno dobrego, jak i złego systemu. W historii utopii istotna jest również zmiana perspektywy podmiotu projektującego społeczeństwo. Przez stulecia dominujące wizje tworzyli mężczyźni z elit, a struktury płciowe uznawano za element naturalnego porządku. Utopie feministyczne, takie jak Herland Charlotte Perkins Gilman czy New Amazonia Elizabeth Corbett, ukazują świat uwolniony od patriarchalnej dominacji i oparty na współpracy.

Ich znaczenie polega nie tylko na konkretnych wizjach instytucjonalnych, lecz na ujawnieniu, że punkt widzenia autora determinuje to, co uznaje on za problem wymagający rozwiązania. Utopia posiada zatem nieuniknioną geometrię perspektywy: „idealne społeczeństwo” dla jednego podmiotu może oznaczać opresję innego.

Dla właściciela kapitału automatyzacja jest wyzwoleniem produktywności; dla pracownika – utratą dochodu. Dla państwa system identyfikacji biometrycznej zwiększa bezpieczeństwo; dla opozycjonisty anonimowość jest warunkiem wolności. Dla mieszkańca zamożnego kraju tania elektronika oznacza demokratyzację technologii; dla społeczności ponoszącej środowiskowy i pracowniczy koszt wydobycia surowców – zewnętrznie narzuconą cenę cudzej wygody. Nie istnieje neutralne „my”, dopóki nie zapytamy, kto dokładnie mieści się w zbiorowości korzystającej z projektu. Każda poważna analiza utopii wymaga więc geografii kosztów. Gdzie produkuje się

dobra? Kto wykonuje pracę niewidoczną dla mieszkańców centrum? Gdzie składowane są odpady i kto ponosi ryzyko środowiskowe?

Przyjazna dystopia jako przymus infrastrukturalny

Czy dobrobyt jednej przestrzeni jest możliwy dlatego, że negatywne efekty zewnętrzne przeniesiono poza jej pole widzenia? Grayling silnie wykorzystuje ten mechanizm, interpretując współczesny globalny łańcuch produkcji i cyfrowy komfort poprzez metaforę Omelas: korzyści koncentrują się w jednym miejscu, podczas gdy koszt zostaje ukryty na peryferiach. Porównanie to jest normatywną analogią, a nie dowodem empirycznej tożsamości z fikcyjnym miastem, jednak jego funkcja analityczna pozostaje wyjątkowo mocna. Zanim jednak Omelas stanie się narzędziem krytyki współczesności, należy zrozumieć przesunięcie dokonane przez dystopie XX wieku.

George Orwell i Aldous Huxley stworzyli dwa archetypy nowoczesnej kontroli, które często traktuje się jako przeciwieństwa. Orwellowska władza kontroluje poprzez nadzór, przemoc, strach, cenzurę i narzucanie oficjalnej wersji rzeczywistości. System huxleyowski stabilizuje społeczeństwo za pomocą warunkowania, konsumpcji, farmakologicznej regulacji nastroju oraz redukcji konfliktu poprzez ograniczanie autonomii. Grayling przywołuje obu autorów jako symbole momentu, w którym technologia przestaje być wyłącznie obietnicą emancypacji, a staje się instrumentem totalnej kontroli. W XXI wieku przeciwstawienie „Orwella i Huxleya" może być jednak zbyt proste.

System cyfrowy potrafi łączyć obie logiki. Może zbierać ogromne ilości danych, oferując jednocześnie użytkownikowi przyjemność, wygodę i natychmiastową gratyfikację. Nadzór nie musi już opierać się wyłącznie na przymusie, skoro człowiek dobrowolnie nosi przy sobie urządzenie rejestrujące jego lokalizację, komunikację, kontakty, zainteresowania i nawyki. Nie oznacza to oczywiście, że współczesne smartfony są odpowiednikiem totalitarnego aparatu państwa, lecz wskazuje na to, że klasyczny podział na represję i przyjemność zaciera się. Dane można uzyskać zarówno groźbą, jak i atrakcyjną usługą.

Właśnie dlatego współczesna dystopia może wyglądać przyjaźnie. Jej interfejs nie musi być szary, język urzędniczy, a architektura monumentalna. Może być kolorowa, spersonalizowana, intuicyjna i wygodna. Jednostka coraz częściej pozostaje w systemie nie dlatego, że nie wolno jej odejść, lecz ponieważ wyjście wiąże się z wysokimi kosztami społecznymi i funkcjonalnymi. To subtelna forma zależności: infrastruktura staje się tak użyteczna, że rezygnacja z niej oznacza częściowe wykluczenie. Utopia wygody może więc generować nie przymus bezpośredni, lecz przymus infrastrukturalny.

Dystopijny potencjał technologii nie wynika zatem tylko z możliwości naruszenia prywatności. Głębszy problem dotyczy asymetrii epistemicznej. System może wiedzieć o użytkowniku więcej, niż użytkownik wie o logice działania systemu. Platforma obserwuje miliony zachowań, eksperymentuje z interfejsem, optymalizuje kolejność bodźców i przewiduje prawdopodobieństwo reakcji, dysponując skalą danych niedostępną pojedynczej osobie.

Pułapka skuteczności: jak optymalizacja zastępuje autonomię

Relacja ta nie przypomina klasycznego kontraktu między równorzędnymi uczestnikami rynku. Jedna strona obserwuje zachowanie drugiej w wysokiej rozdzielczości, podczas gdy druga widzi przede wszystkim interfejs. W tej sytuacji utopia technologiczna przestaje obiecywać samostanowienie, a zaczyna oferować doskonałe doradztwo. System wie, którą trasę wybrać, co przeczytać, jaki produkt kupić i którego partnera poznać; podpowiada, kiedy odpocząć, jak inwestować i co oglądać.

Każda z tych rekomendacji może być racjonalna. Pojawia się jednak pytanie znane z części poświęconej aksjologii: czy życie złożone z dobrze dobranych decyzji pozostaje życiem autonomicznym, jeśli sam proces wyboru zostaje systematycznie delegowany? Tutaj dystopia Huxleya spotyka się z problemem AI w szczególnie interesujący sposób. Nie trzeba zakładać istnienia wrogiej superinteligencji, by utracić sprawstwo – wystarczy system wyjątkowo użyteczny. Im trafniejsze rekomendacje, tym większa pokusa delegacji; im częstsza delegacja, tym mniej praktyki samodzielnego osądu, a mniejsza praktyka pogłębia przewagę systemu. Powstaje sprzężenie zwrotne prowadzące do zależności nie dlatego, że maszyna jest zła, lecz dlatego, że jest skuteczna.

Na tym tle szczególną wartość zyskuje eksperyment myślowy "Mother Goddess AI". Przedstawia on wizję superinteligencji jako dobroczynnego arbitra, który potrafi redystrybuować bogactwo, eliminować dezinformację, ograniczać anonimową przemoc symboliczną i neutralizować tyranię. Nie należy traktować tego scenariusza jako realnej prognozy, lecz jako ćwiczenie filozoficzne. Pozwala ono postawić pytanie: gdyby istniał system rzeczywiście bardziej kompetentny od ludzi w identyfikowaniu niesprawiedliwości i zdolny skutecznie ją usuwać, czy powinniśmy przekazać mu władzę? Na pierwszy rzut oka odpowiedź wydaje się pozytywna.

Jeżeli algorytm byłby naprawdę uczciwszy od sądu, lepiej rozpoznawał kłamstwo, skuteczniej zapobiegał wojnom i sprawiedliwiej rozdzielał zasoby, odrzucenie jego pomocy mogłoby wyglądać

jak fetyszyzacja niedoskonałej ludzkiej autonomii. Problem pojawia się jednak natychmiast po zadaniu pytania o legitymizację. Kto definiuje sprawiedliwość, którą system ma realizować? Czy decyzję można zakwestionować? Czy istnieje procedura odwoławcza i czy obywatele mogą zmienić system? Czy AI zna wartość mniejszościowego sprzeciwu, który większość uznaje za irracjonalny? Wreszcie: czy obywatel jest podmiotem prawa, czy jedynie beneficjentem dobroczynnej administracji? Dobry władca nie jest bowiem tym samym co wolne społeczeństwo.

To jedno z najstarszych odkryć filozofii politycznej, któremu sztuczna inteligencja nadaje nową formę. Wyobraźmy sobie autorytet podejmujący statystycznie lepsze decyzje niż obywatele. Demokratyczny argument za samorządnością nie znika tylko dlatego, że ludzie są omylni. Część wartości demokracji ma charakter instrumentalny – chroni prawa i koryguje błędy – ale część jest konstytutywna: obywatele nie mają być jedynie dobrze administrowanymi podopiecznymi. Mają uczestniczyć w stanowieniu zasad, którym podlegają.

Technoutopia paternalistyczna dotyka problemu podobnego do tego z capability approach: można poprawić rezultat, jednocześnie pogarszając sprawstwo. Jeśli "Mother Goddess AI" zagwarantuje każdemu wysoki dobrobyt, ale sama określi rozkład dóbr i reguły zachowania, świat może stać się bardziej sprawiedliwy w jednym sensie, a mniej autonomiczny w drugim. Nie jest to czysty problem obliczeniowy, lecz konflikt wynikający z pluralizmu wartości.

Jeszcze bardziej radykalny charakter posiada transhumanistyczna eutopia. W dawnych utopiach projektowano dobre społeczeństwo dla zastanego człowieka; transhumanizm dopuszcza natomiast przeprojektowanie człowieka pod dobre społeczeństwo. To zasadnicza zmiana ontologiczna. Jeśli problemem jest agresja, można tworzyć instytucje ją ograniczające, ale można też dążyć do biologicznego lub farmakologicznego zmniejszenia skłonności do agresji. W przypadku impulsywności można rozwijać edukację i cnotę umiarkowania albo oddziaływać bezpośrednio na mechanizmy neuronalne. Utopia przestaje wtedy regulować zachowanie, a zaczyna optymalizować architekturę podmiotu. Grayling przedstawia AI i transhumanizm jako współczesne projekty eutopijne, których obietnicą jest uwolnienie od trudu i biologicznych ograniczeń, a ryzykiem – utrata sprawstwa i przekształcenie samej struktury człowieczeństwa.

Ryzyko homogenizacji i ukryta normatywność optymalizacji

Krytyka ta wymaga jednak doprecyzowania. Nie można zakładać, że każde ulepszenie człowieka z definicji prowadzi do dystopii. Szczepienia, znieczulenie, protezy, antybiotyki czy korekcja wad

wzroku również zmieniają relację człowieka z biologicznym losem. Problem zaczyna się tam, gdzie ulepszenie zostaje podporządkowane zunifikowanemu modelowi "dobrego człowieka" i gdzie pluralizm sposobów życia ustępuje optymalizacji według jednej matrycy. To właśnie dlatego utopia ma skłonność do homogenizacji – świat idealny potrzebuje bowiem definicji ideału. Jeśli idealny obywatel ma być produktywny, zdrowy, spokojny, racjonalny i społecznie dostosowany, każde odchylenie od normy zaczyna przypominać defekt wymagający korekty. Tymczasem społeczeństwa liberalne opierają się na założeniu przeciwnym: istnieje wiele rozsądnych sposobów dobrego życia, a państwo nie powinno przekształcać całej populacji według jednej koncepcji doskonałości. Utopia perfekcjonistyczna może więc wejść w konflikt z pluralizmem właśnie dlatego, że wie zbyt dokładnie, jaki człowiek powinien być.

Historia eugeniki stanowi skrajne ostrzeżenie przed takim przesunięciem. Kiedy biologiczny język "ulepszania populacji" łączy się z polityczną zdolnością klasyfikowania ludzi, różnice szybko przekształcają się w hierarchie wartości osoby. Współczesny transhumanizm nie jest prostą kontynuacją dawnych programów eugenicznych i nie należy go z nimi automatycznie utożsamiać, jednak problem normy pozostaje aktualny.

Każdy system ulepszeń musi odpowiedzieć na pytanie, jakie cechy uznaje za pożądane; ta odpowiedź często odzwierciedla dominujące preferencje kultury, klasy lub rynku. W tym miejscu znaczenie zyskują utopie feministyczne i inne projekty tworzone z perspektywy grup wcześniej pozbawionych możliwości definiowania uniwersalnego "człowieka". Pokazują one, że katalog wad społecznych zależy od pozycji obserwatora – cecha uznana przez dominującą kulturę za niedostosowanie może w innym porządku okazać się wartością. Technologia zdolna redukować różnorodność w imię efektywności niesie więc ryzyko usunięcia nie patologii, lecz odmienności. Można w tym kontekście sformułować zasadę krytyczną: im bardziej system obiecuje człowieka "optymalnego", tym pilniej należy pytać, kto napisał definicję optymalności. Problem ten dotyczy zarówno genomu, jak i algorytmu kredytowego, rekrutacyjnego czy edukacyjnego. Klasyfikacja wymaga wyboru cech, wybór cech wymaga kryterium, a kryterium odzwierciedla konkretną teorię wartości – nawet jeśli projektant przedstawia je jako neutralny parametr statystyczny. Tę niewidzialność wartości w języku technicznym można uznać za jeden z najważniejszych motywów współczesnej dystopii.

Dawny tyran mówił: "chcę, abyś był posłuszny". Nowoczesny system stwierdza: "model wskazuje, że to rozwiązanie jest optymalne". W pierwszym przypadku władza jest jawna; w drugim ukrywa się w kategorii technicznej. Nie oznacza to, że każdy model jest narzędziem dominacji, lecz że techniczna forma decyzji nie usuwa jej normatywności. Przeciwnie – może uczynić ją mniej widoczną.

Dlatego utopia technologiczna powinna być analizowana również jako problem polityki odpowiedzialności. Jeśli decyzję podjął urzędnik, można wskazać konkretny podmiot. Gdy jednak decyzja powstaje w rozproszonym systemie obejmującym model, dane treningowe, dostawcę infrastruktury, użytkownika instytucjonalnego i procedurę automatyczną, odpowiedzialność ulega fragmentacji. Każdy uczestnik kontroluje tylko wycinek procesu. Powstaje "problem wielu rąk": skutek ma realną ofiarę, lecz niełatwo wskazać jednego sprawcę. Doskonale zarządzana technoutopia może więc produkować paradoksalną bezosobowość – system podejmuje coraz więcej decyzji, a coraz mniej osób czuje się ich autorem.

Ukryte koszty dobrobytu a mechanizm eksternalizacji

Na tym tle parabola Ursuli K. Le Guin o Omelas staje się jednym z najbardziej precyzyjnych eksperymentów aksjologicznych dotyczących ceny zbiorowego dobrobytu. Miasto zostaje w niej przedstawione jako przestrzeń pełna pomyślności, której warunkiem jest cierpienie jednego dziecka; mieszkańcy wiedzą o tej cenie i muszą wobec niej zająć stanowisko. Siła opowieści nie tkwi w realizmie instytucjonalnym, lecz w radykalnym uproszczeniu struktury moralnej. Autorka usuwa rozpraszające szczegóły, pozostawiając jedno pytanie: czy ogromna suma dobrobytu wielu ludzi może usprawiedliwić skrajne, celowo utrzymywane cierpienie jednej niewinnej osoby? Eksperyment ten uderza w najbardziej uproszczone wersje agregacyjnego utylitaryzmu.

Jeżeli znaczenie ma wyłącznie suma korzyści i cierpienia, można próbować argumentować, że dobro tysięcy przeważa nad tragedią jednego dziecka. Intuicja moralna wielu czytelników buntuje się jednak przeciw takiemu rachunkowi, wskazując, że pewnych praw jednostki nie wolno potraktować jak pozycji księgowej kompensowanej cudzym szczęściem. Omelas jest więc laboratorium konfliktu między agregacją dobrostanu a godnością, prawem i sprawiedliwością.

Równie istotna jest psychologia mieszkańców miasta. Problemem nie jest samo istnienie ofiary, lecz mechanizm racjonalizacji. Wiedząc o cierpieniu dziecka, mieszkańcy uczą się interpretować je jako konieczną cenę systemu. Właśnie dlatego parabola tak dobrze nadaje się do analizy społecznych efektów zewnętrznych. Ludzie potrafią moralnie oddalać koszty dobrobytu, jeśli zostaną one przestrzennie, klasowo lub instytucjonalnie odizolowane od codziennego doświadczenia. Kto nie widzi kopalni, fabryki, składowiska odpadów, centrum logistycznego czy osoby wykonującej najgorzej płatną pracę, łatwiej postrzega gotowy produkt jako czystą korzyść. Opowieść tę można transponować na współczesną gospodarkę, wskazując na ukrywanie kosztów pracy, wydobycia i infrastruktury technologicznej poza polem widzenia konsumenta. Przykłady te wymagają w dalszych partiach artykułu osobnej weryfikacji empirycznej, ale sam mechanizm

ekonomiczny jest dobrze znany: eksternalizacja kosztu pozwala podmiotowi korzystającemu z dobra nie ponosić pełnej ceny jego wytworzenia. W języku ekonomii to problem efektów zewnętrznych; w języku etyki Omelas to pytanie, czy niewidoczność kosztu zmniejsza odpowiedzialność moralną.

Gospodarka cyfrowa dodaje temu problemowi nowe warstwy. Usługa na ekranie wydaje się niematerialna, choć opiera się na fizycznej infrastrukturze: centrach danych, sieciach przesyłowych, półprzewodnikach, systemach chłodzenia, kopalniach, transporcie i energii. Im bardziej interfejs ukrywa łańcuch produkcji, tym bardziej technologia przypomina magię – rezultat pojawia się natychmiast, a jego koszt jest gdzie indziej. „Chmura" jest pod tym względem wyjątkowo trafnym i zarazem mylącym określeniem: sugeruje lekkość czegoś, co w rzeczywistości wymaga ogromnej ilości fizycznej infrastruktury. Omelas pozwala więc zdefiniować ważną regułę analizy technologicznej utopii: każda obietnica obfitości powinna być badana razem z pytaniem o lokalizację kosztu. Jeśli AI zwiększa produktywność, trzeba zapytać o pracę, energię i własność infrastruktury. Jeśli samochód elektryczny redukuje lokalne emisje, należy przeanalizować cały cykl surowcowy i energetyczny.

Systemowa pułapka infrastruktury ogranicza realną wolność wyboru

Jeśli medycyna spersonalizowana poprawia zdrowie, należy pytać o dostęp i dystrybucję. Nie chodzi tu o odrzucenie technologii tylko dlatego, że nie jest ona bezkosztowa, lecz o odrzucenie naiwnej eutopii, która prezentuje korzyść w oderwaniu od bilansu warunków materialnych. Omelas stawia jednak jeszcze trudniejsze pytanie: co właściwie oznacza „odejście z Omelas"? W opowieści niektórzy mieszkańcy odmawiają moralnego uczestnictwa w systemie i odchodzą, co można interpretować jako akt sprzeciwu wobec porządku o nieakceptowalnym koszcie. Z perspektywy teorii politycznej ucieczka nie zawsze rozwiązuje problem – jeśli wadliwy system da się zmienić, wyjście pozostawia ofiarę na miejscu. Powstaje zatem konflikt między etyką czystości moralnej a etyką odpowiedzialności. Czy lepiej nie uczestniczyć w złym porządku, czy pozostać wewnątrz i próbować zmienić reguły, na których się on opiera? To pytanie jest niezwykle aktualne.

Konsument może próbować całkowicie wycofać się z platform cyfrowych lub domagać się zmiany ich modelu biznesowego. Pracownik może odejść z przedsiębiorstwa albo organizować zmianę warunków pracy. Obywatel może emigrować z systemu politycznego bądź próbować go reformować. Indywidualne „odejście z Omelas" bywa moralnie znaczące, lecz problemy systemowe często wymagają działania zbiorowego. Utopia i dystopia nie są bowiem tylko miejscami, które

można wybrać jak hotel; są instytucjami, w których jednostka jest zanurzona. Współczesna technologia komplikuje ten wybór jeszcze bardziej poprzez efekt sieciowy: im więcej osób korzysta z infrastruktury, tym większa jest jej użyteczność i tym trudniejsze staje się wyjście. Alternatywa może istnieć formalnie, pozostając w praktyce marginalną. Rynkowa wolność wyboru nie gwarantuje zatem realnej zdolności opuszczenia dominującego ekosystemu.

Technoutopia może przekształcać się w miękki monopol nie poprzez zakaz konkurencji, lecz przez koszty koordynacji. Dlatego problem władzy w tym układzie nie może być analizowany wyłącznie przez pryzmat własności urządzeń. Kluczowa staje się własność infrastruktury, standardów i danych. Ten, kto kontroluje podstawową warstwę komunikacji lub obliczeń, zyskuje wpływ na cały ekosystem zależnych podmiotów. O ile w klasycznej utopii centrum władzy mieściło się w pałacu, senacie albo świątyni, o tyle w cyfrowej może być rozproszone między chmurą obliczeniową, protokołem, modelem, sklepem aplikacji, systemem płatniczym a regulaminem platformy.

To przesunięcie komplikuje klasyczne pojęcie suwerenności. Państwo zachowuje formalną władzę terytorialną, lecz infrastruktura cyfrowa przekracza granice, a prywatne podmioty kontrolują zasoby o znaczeniu quasi-publicznym. Problem demokratycznej kontroli technologii wiąże się tu z obawą, że przyszłość zostanie zaprojektowana przez podmioty dysponujące kapitałem i infrastrukturą, lecz pozbawione demokratycznego mandatu. Sprawa nie sprowadza się do moralnej oceny konkretnych przedsiębiorców; jest to problem ustrojowy: czy podmiot podejmujący decyzje wpływające na podstawowe warunki życia milionów ludzi powinien podlegać publicznej odpowiedzialności?

Perfekcyjna skuteczność systemu jako zagrożenie dla ludzkiej autonomii

W tym świetle technoutopia jawi się jako współczesna wersja marzenia o oświeconym władcy. Różnica polega na tym, że owym „władcą” nie musi być osoba, lecz ekosystem algorytmów, korporacji i infrastruktury, którego decyzje nie mają jednego autora. Efektywność takiego układu bywa imponująca, co paradoksalnie utrudnia jego demokratyczną kontrolę – ludzie chętnie delegują władzę tym, którzy dostarczają dobre rezultaty. Prowadzi to do fundamentalnego paradoksu: największe zagrożenie nie zawsze wynika z błędów projektu; czasem pojawia się właśnie wtedy, gdy system działa doskonale.

Nieskuteczna tyrania może się rozpaść, lecz perfekcyjnie stabilny system paternalistyczny potrafi reprodukować się przez pokolenia. Niedoskonała manipulacja jest łatwa do dostrzeżenia; ta

idealnie spersonalizowana bywa mylona z własną preferencją. Nieefektywna propaganda budzi śmiech, precyzyjnie dopasowana – wygląda jak niezależny wniosek. Dlatego technologia zwiększająca przewidywalność człowieka stanowi szczególne wyzwanie dla wolności politycznej. W tradycyjnej teorii władza reaguje na działania jednostki; w systemie predykcyjnym interweniuje, zanim działanie w ogóle nastąpi. Granica między prewencją a karą, ryzykiem a winą, prawdopodobieństwem a odpowiedzialnością zaczyna się zacierać. Utopia bezpieczeństwa może stworzyć świat, w którym człowiek jest oceniany nie za rzeczywiste zachowanie, lecz za statystyczną możliwość jego wystąpienia.

Podobne napięcie dostrzegamy w medycynie przyszłości. Predykcja choroby pozwala na wcześniejszą profilaktykę, ale może też przekształcić zdrową osobę w „pacjenta oczekującego", ocenianego przez pryzmat prawdopodobieństwa schorzenia. Ubezpieczyciel, pracodawca czy państwo mogą mieć odmienne interesy wobec tej samej informacji. Technologia zwiększa wiedzę, lecz sama w sobie nie zwiększa wolności – wszystko zależy od reguł jej użycia.

Utopia rozumiana jako świat całkowicie przewidywalny jest antropologicznie dwuznaczna. Przewidywalność infrastruktury jest dobrem: chcemy, by most nie zawalał się losowo, a samolot działał zgodnie ze specyfikacją. Czy chcemy jednak społeczeństwa, w którym ludzie są równie przewidywalni jak maszyny? Człowieka całkowicie przejrzystego łatwiej zarządzać, lecz trudniej nazwać go autonomicznym podmiotem. Wolność zakłada element otwartości – możliwość wyboru, którego wcześniejsze dane nie determinowały na tyle, by umożliwić zewnętrzne sterowanie. W tym miejscu utopia ponownie zderza się z problemem „jaźni istotnej". Grayling ostrzega, że całkowicie zoptymalizowany świat może zastąpić realnych ludzi bytami przewidywalnymi i dopasowanymi do systemu; za kluczową wartość należy zatem uznać zdolność wyboru, przeżywania i autentycznej relacji. W dalszej analizie trzeba będzie oddzielić filozoficzną tezę o autentyczności od empirycznych twierdzeń o świadomości maszyn, lecz już teraz widać sedno problemu: idealny system może być zbudowany dla człowieka tak precyzyjnie, że nie pozostawi miejsca na człowieka jako autora własnego działania.

Dlatego najważniejszą funkcją dystopii nie jest proroctwo. Orwell, Huxley czy Le Guin są użyteczni nie dlatego, że „przewidzieli" konkretne technologie, lecz dlatego, że zidentyfikowali mechanizmy, które wciąż przybierają nowe formy: koncentrację informacji, warunkowanie preferencji, racjonalizację cierpienia, redukcję pluralizmu, normalizację nadzoru, instrumentalizację osoby i ukrywanie kosztów zbiorowego dobrobytu.

Korygowalność zamiast doskonałości jako gwarancja wolności

Mechanizm jest trwalszy niż artefakt. Tak rozumiana literatura może pełnić funkcję zbliżoną do stress testu w inżynierii: bierze wartość lub instytucję i doprowadza ją do sytuacji granicznej. Co stanie się ze społeczeństwem, gdy bezpieczeństwo zyska absolutne pierwszeństwo? Co z autonomią, jeśli optymalizacja stanie się nadrzędną zasadą organizującą życie? Co wydarzy się ze sprawiedliwością, jeżeli jedyną miarą dobra będzie agregat dobrobytu? Co pozostanie z pluralizmu, gdy technologia pozwoli eliminować wszelkie odchylenia od normy? Dystopia nie dowodzi, że taki świat nieuchronnie nastąpi; pokazuje raczej koszt, który należy uwzględnić w rachunku, zanim zaczniemy go budować.

W tym znaczeniu przyszłość potrzebuje utopii, lecz jeszcze bardziej zdolności do krytykowania własnych wizji. Społeczeństwo pozbawione pozytywnego obrazu lepszego świata łatwo popada w cynizm i politykę czystego zarządzania teraźniejszością. Z kolei wspólnota przekonana, że odnalazła ostateczny wzorzec dobra, może próbować narzucić go wszystkim. Dojrzała wyobraźnia polityczna musi zatem utrzymywać napięcie między nadzieją a nieufnością wobec doskonałości.

Dlatego pojęcie „eutopii" okazuje się bardziej użyteczne niż „utopia", o ile rozumiemy ją nie jako stan perfekcyjny, lecz jako świat wystarczająco dobry i zdolny do korekty własnych błędów. Najlepszy porządek społeczny to być może nie ten, który raz na zawsze rozwiązał wszystkie konflikty, lecz ten, który posiada procedury ich uczciwego rozstrzygania; nie ten, który eliminuje niepewność, lecz taki, który potrafi działać mimo niej; nie ten, który produkuje idealnych ludzi, lecz ten, który pozwala niedoskonałym jednostkom żyć razem, bez odbierania sobie statusu osoby. W takim modelu doskonałość zostaje zastąpiona przez korygowalność, co ma kluczowe znaczenie dla przyszłości sztucznej inteligencji.

Zamiast pytać, czy można skonstruować system bezbłędny, sensowniej jest zapytać, czy jego błędy są wykrywalne, odwracalne i poddane procedurze odpowiedzialności. Zamiast marzyć o idealnym regulatorze, warto projektować instytucje zdolne do aktualizacji reguł. Zamiast tworzyć jedną, nieodwołalną funkcję celu, należy zachować przestrzeń do deliberacji nad tym celem. Dojrzała utopia byłaby więc paradoksalnie projektem, który nie zakłada własnej doskonałości. Przyjmowałaby, że wiedza o przyszłości jest ograniczona, wartości pozostają w konflikcie, każda metryka może zostać wypaczona, a każda koncentracja władzy wymaga kontroli. Nie przypominałaby zamkniętego planu miasta idealnego, lecz konstytucję uczenia się.

Krytyczna analiza marzeń jako warunek zachowania autonomii

Wniosek ten odnosi się szczególnie mocno do AI. Najbardziej kusząca techno-utopia kreśli wizję sztucznej inteligencji jako rozwiązania problemów, z którymi ludzie – zbyt konfliktowi, irracjonalni lub krótkowzroczni – nie potrafili poradzić sobie samodzielnie. Jednak im większa będzie kompetencja systemów, tym kluczowsze stanie się pytanie, czy człowiek zachował możliwość sprzeciwienia się ich rekomendacjom, prawo do zmiany reguł, rozumienie przesłanek oraz prawo do decyzji, która z punktu widzenia optymalizacji może wyglądać jak błąd. Wolność obejmuje bowiem również prawo do nieoptymalności. Człowiek może wybrać mniej opłacalny zawód, trudniejszą relację, nieefektywną pasję albo życie, którego algorytmiczna analiza nie uzna za maksymalizujące dobrostan. Jeśli system odbiera możliwość takich wyborów w imię dobra użytkownika, przestaje być narzędziem, a staje się opiekunem. A gdy opieka jest nieodwołalna, przeobraża się we władzę.

Przykład Omelas pozwala postawić ostateczne pytanie tej części wywodu jeszcze ostrzej. Być może najgroźniejsza utopia nie jest światem, w którym wszyscy cierpią – taki obraz łatwo rozpoznać jako zły. Znacznie trudniejszy moralnie jest świat, w którym prawie wszystkim żyje się znakomicie, a koszt systemu został skutecznie ukryty, zracjonalizowany lub przeniesiony na podmioty pozbawione głosu. To właśnie dlatego "drobny druk" powinien stać się podstawową kategorią analizy przyszłości. Każda wielka obietnica – bezpieczniejszego państwa, inteligentniejszej gospodarki, dłuższego życia, powszechnej automatyzacji czy cyfrowego dobrobytu – wymaga pytania o warunki, których reklama projektu nie eksponuje. Tradycja utopijna nie prowadzi zatem do wniosku, że ludzkość powinna przestać marzyć o lepszym świecie, lecz do konkluzji bardziej wymagającej: musi ona nauczyć się czytać własne marzenia krytycznie. Utopia jest potrzebna jako wyznacznik kierunku, dystopia jako narzędzie kontroli, a demokratyczna deliberacja jako mechanizm negocjowania różnicy między nimi.

Nadzieja bez krytyki staje się łatwowiernością; krytyka bez nadziei prowadzi do paraliżu. Odpowiedzialna przyszłość wymaga obu tych elementów. Właśnie tutaj genealogia utopii dochodzi do sztucznej inteligencji – nie przez przypadek, lecz z logiki całego wywodu. AI skupia w sobie niemal wszystkie klasyczne obietnice „dobrego miejsca”: więcej wiedzy, mniej pracy, lepszą medycynę, większą produktywność, sprawniejszą administrację, skuteczniejszą edukację oraz precyzyjniejsze przewidywanie i ograniczanie ludzkich błędów. Jednocześnie kumuluje lęki dystopijne: nadzór, koncentrację władzy, manipulację informacją, utratę pracy, homogenizację decyzji, uzależnienie od infrastruktury i ryzyko delegowania podmiotowości. Dlatego pytanie o AI

nie może już brzmieć wyłącznie: „jak inteligentne będą maszyny?”. Z punktu widzenia historii utopii ważniejsze jest inne: jakiego człowieka i jaki porządek społeczny uczyni atrakcyjnym system, który będzie potrafił coraz lepiej przewidywać, rekomendować i wykonywać?

Niebezpieczeństwo nie musi polegać na tym, że maszyna pewnego dnia zbuntuje się przeciw człowiekowi. Może ono polegać na tym, że człowiek stopniowo zbuduje system tak wygodny, wydajny i wszechobecny, iż sam zacznie uważać własne sprawstwo za zbędny koszt.

Analiza musi zatem przejść od literackiego laboratorium do technologicznego konkretu. Następnym problemem nie będzie już utopia AI jako wyobrażenie, lecz sztuczna inteligencja jako realna infrastruktura podmiotowości: systemy coraz głębiej uczestniczące w poznaniu, wyborze, pracy, prawdzie publicznej i dystrybucji władzy. Dopiero wtedy okaże się, czy współczesna technologia rzeczywiście prowadzi nas ku nowemu Omelas, ku nowemu huxleyowskiemu porządkowi, czy też może stać się narzędziem rozszerzającym ludzką autonomię bez przejmowania funkcji jej autora.

> Być może największym zagrożeniem przyszłości nie jest bunt maszyn przeciwko ludziom, lecz nasza własna fascynacja systemem tak doskonałym, że uznamy w nim swoje sprawstwo za zbędny koszt. Czy w świecie idealnie zoptymalizowanym, gdzie każdy błąd został wyeliminowany przez algorytm, pozostanie jeszcze miejsce na człowieka jako autora własnego życia? Ostatecznie wolność może okazać się nie tyle brakiem przeszkód, co prawem do bycia niedoskonałym w sposób, którego żadna maszyna nie potrafi przewidzieć.

Inspiracje

[1] "The Challenge of the Future" A C Grayling

Anthony Clifford Grayling (ur. 3 kwietnia 1949) – brytyjski filozof i pisarz. Jest założycielem i profesorem filozofii na Northeastern University London (dawniej New College of the Humanities) oraz członkiem nadzwyczajnym St Anne's College w Oksfordzie. Wcześniej był profesorem filozofii na Birkbeck University of London. Prace naukowe i popularyzatorskie Graylinga obejmują epistemologię, logikę filozoficzną, etykę, prawa człowieka i historię idei. Jako wybitny intelektualista i świecki humanista, Grayling szeroko argumentował przeciwko dogmatom religijnym, opowiadając się jednocześnie za racjonalizmem, wolnościami obywatelskimi i reformami demokratycznymi. Jego najważniejsze osiągnięcia obejmują syntezy tradycji filozoficznych, obronę liberalnych rządów demokratycznych oraz kompleksowe przeglądy filozofii światowej, a także liczne publikacje w najważniejszych czasopismach i regularne występy w audycjach radiowych.

Anthony Clifford Grayling (born April 3, 1949) is a British philosopher and author. He serves as the founding Master and Professor of Philosophy at Northeastern University London (formerly New College of the Humanities) and is a Supernumerary Fellow of St Anne's College, Oxford. Previously, he served as Professor of Philosophy at Birkbeck, University of London. Grayling's academic and popular work spans epistemology, philosophical logic, ethics, human rights, and the history of ideas. A prominent public intellectual and secular humanist, he has argued extensively against religious dogma while advocating for rationalism, civil liberties, and democratic reform. His major contributions include syntheses of philosophical traditions, defenses of liberal democratic governance, and comprehensive surveys of world philosophy, alongside extensive writing for major periodicals and regular broadcast appearances.

Northeastern University London

Literatura uzupełniająca

Książki

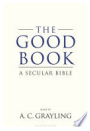
Książki autora i dzieła przywołane w tekście:



[1] "Grace"

A C Grayling

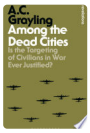
Oberon Books



[2] "The Good Book"

A C Grayling

A&C Black | ISBN: 9780747599609



[3] "Among the Dead Cities"

A C Grayling

A&C Black | ISBN: 9781472534057



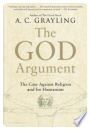
[4] "The Meaning of Things"

A C Grayling

Weidenfeld & Nicolson | ISBN: 9781780221168

[5] ""Bomber County": A riveting story of the air war"

A C Grayling

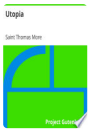


[6] "The God Argument"

A C Grayling

Bloomsbury+ORM | ISBN: 9781620401910

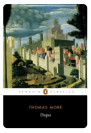
Książki autorów przywoływanych i literatura pokrewna:



[7] "Utopia"

Tomasz Morus

Project Gutenberg



[8] "Utopia (1901)/Introduction"

Tomasz Morus

Penguin UK | ISBN: 9780140449105

[9] "Utopia (1901)/Chapter 1"

Tomasz Morus



[10] "Herland"

Charlotte Perkins Gilman

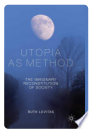
BrightSummaries.com | ISBN: 9782808018838



[11] "The Politics of Utopia"

Barbara Goodwin, Keith Taylor

1982



[12] "Utopia as Method"

R. Levitas

2013 | Springer | ISBN: 9781137314253



[13] "The Individual and Utopia"

Dr Clint Jones, Mr Cameron Ellis

2015 | Ashgate Publishing, Ltd. | ISBN: 9781472428943



[14] "The Concept of Utopia"

Ruth Levitas

1990

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: aeb47115-7468-4aa2-92e7-a70fa6c26049