Głęboka utopia: sens, tożsamość i granice doskonałości

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł podejmuje temat 'głębokiej utopii' w kontekście dynamicznego rozwoju sztucznej inteligencji i automatyzacji. Autor analizuje zagrożenie apatią ontologiczną oraz entropią praktycznego rozumu, które mogą pojawić się w świecie absolutnej obfitości. Kluczowym elementem rozważań jest koncepcja celów niewymiennych – działań, których wartość wynika z samego faktu ich osobistego wykonywania, a nie tylko z końcowego efektu. Tekst odwołuje się do myśli Nicka Bostroma, badając granice ludzkiej tożsamości i sprawczości. Poprzez metaforę 'oporu świata', artykuł rekonstruuje sens i wolę jako niezbędne składniki ludzkiego dobrostanu. To filozoficzna podróż w stronę postludzkiej przyszłości, w której zachowanie agent-relatywności staje się fundamentem przetrwania sensu życia w erze post-pracy, chroniąc nas przed utratą tożsamości w obliczu doskonałych maszyn.

This article explores the topic of 'deep utopia' in the context of the dynamic development of artificial intelligence and automation. The author analyzes the threat of ontological apathy and the entropy of practical reason, which may emerge in a world of absolute abundance. A key element of this discussion is the concept of non-exchangeable goals – actions whose value derives from the very fact of their personal performance, not just from the final outcome. The text draws on the thought of Nick Bostrom, exploring the limits of human identity and agency. Through the metaphor of 'world resistance,' the article reconstructs meaning and will as essential components of human well-being. It is a philosophical journey toward a posthuman future in which the preservation of agent-relativity becomes the foundation for the survival of meaning in the post-work era, protecting us from the loss of identity in the face of perfect machines.

W dobie postępującej automatyzacji i wizji utopijnego świata, w którym maszyny wyręczają człowieka we wszelkich czynnościach, kluczowe staje się pytanie o źródło sensu i wartości ludzkiego istnienia. Artykuł analizuje koncepcję agent-relatywności, czyli idei, że pewne cele i działania zyskują wartość wyłącznie dzięki tożsamości wykonawcy, a nie jedynie ze względu na

rezultat. Te działania, związane z relacjami, narracją i osobistym zaangażowaniem, stają się niezastąpione w świecie, w którym maszyny przejmują zadania agent-neutralne. Autor argumentuje, że utopia, by pozostać zamieszkałą i sensowną, musi celowo zachować obszary, w których ludzka sprawczość jest niezastąpiona, a tożsamość podmiotu odgrywa kluczową rolę. W przeciwnym razie, grozi nam apatia ontologiczna – utrata woli i pragnienia, które definiują człowieczeństwo.

SŁOWA KLUCZOWE

Głęboka utopia | agent-relatywność | agent-neutralność | cele niewymienne | Nick Bostrom | apatia ontologiczna | entropia praktycznego rozumu | sprawczość | automatyzacja | sens życia | wola | gradient wartości | opór świata | postludzka filozofia | tożsamość

Agent-relatywność: fundament sensu w utopii

Zacznijmy od przypowieści, która odsłania naturę problemu ostrzej niż jakikolwiek abstrakcyjny argument. W odległej przyszłości istniała cywilizacja, w której maszyny o doskonałej precyzji i nieograniczonej lojalności mogły wykonać każde zadanie. Pewnego dnia kobieta – nazwijmy ją Lira – postanowiła własnoręcznie zrobić naszyjnik dla przyjaciółki. Maszyny natychmiast zaoferowały pomoc: „Możemy stworzyć dzieło nieporównanie piękniejsze, idealnie dopasowane, trwalsze". Lira odpowiedziała: „Nie chodzi o piękno. Chodzi o to, że to ja chcę go zrobić". Maszyny kontynuowały: „Ale po co marnować czas, skoro możesz osiągnąć lepszy rezultat bez wysiłku?". Lira odparła krótko: „Rezultat nie jest celem. Celem jest to, że ja jestem przyczyną".

W tej jednej ripoście zawiera się cała ontologia agent-relatywności: istnieją cele, których wartość nie zależy od jakości wykonania, lecz od tożsamości wykonawcy. Agent-relatywność to strukturalna własność relacji, w której osoba staje się niezastąpiona, a sens nie tkwi w przedmiocie działania, ale w podmiocie. W naszym przedutopijnym świecie była to oczywistość – rodzicielstwo, troska, solidarność czy odpowiedzialność polityczna zakładały nieredukowalność konkretnej osoby. W utopii jednak normą staje się agent-neutralność, czyli przekonanie, że liczy się tylko optymalny wynik, a maszyny wykonają wszystko lepiej, szybciej i bezbłędnie.

Tu właśnie rodzi się fundamentalny problem. Jeśli maszyna zrobi wszystko, a człowiek nie musi zrobić niczego, co wymaga jego osobistej obecności, to człowiek staje się opcjonalny w świecie, który sam stworzył. Nick Bostrom wskazuje, że w takiej utopii zadania agent-neutralne zostaną zredukowane do zera.

Maszyny przejmą całą sferę tradycyjnej sprawczości – naprawę, opiekę, produkcję, analizę, a nawet nauczanie i terapię – czyniąc ludzki wysiłek redundantnym. Co zatem pozostaje? Pozostają wyłącznie cele, które z definicji wymagają człowieka. Nazwijmy je celami niewymiennymi, aby podkreślić, że ich wartość nie podlega wymianie ani delegacji.

Cele te mają cztery nierozerwalne właściwości. Po pierwsze, ich wartość zależy od osoby, a nie od wyniku. Po drugie, ich treść jest konstruowana przez relację, a nie przez funkcję. Po trzecie, ich znaczenia nie da się zoptymalizować. I wreszcie po czwarte, ich sens jest wytwarzany przez narrację, nie przez użyteczność. W tej perspektywie agent-relatywność nie jest archaiczną pozostałością epoki biologicznej, lecz najbardziej fundamentalnym źródłem sensu w przyszłości. Spróbujmy dokonać formalnej rekonstrukcji tej struktury.

Niech AR oznacza działanie, którego wartość wynika z faktu, że wykonuje je konkretna osoba X. Niech AN oznacza działanie o wartości niezależnej od wykonawcy. Niech F oznacza zadanie w pełni automatyzowalne, a S – działanie generujące sens. Przyjmujemy następujące zależności: jeśli działanie jest agent-neutralne (AN) i da się je zautomatyzować (F), to nie generuje sensu (nie S). Jeśli natomiast jest agent-relatywne (AR), to może generować sens (możliwe S). W utopii większość zadań AN staje się zbędna, dlatego sens (S) zależy niemal wyłącznie od działań typu AR. Oznacza to, że w świecie absolutnej obfitości sens staje się funkcją konkretnego „ja”, a nie funkcją świata.

W tym miejscu trzeba odwołać się do rozumowania Bostroma: „Jeżeli utopia ma pozostać zamieszkana, musi chronić miejsca, w których tylko człowiek jest wystarczająco ludzki, by być sprawcą”.

Natychmiastowa doskonałość anihiluje podmiot

Wyobraźmy sobie cywilizację, w której wszystkie dobra, usługi i działania mogą być wykonywane przez maszyny o doskonałej precyzji i nieograniczonej lojalności. Pewnego dnia kobieta – nazwijmy ją Lira – postanowiła podarować przyjaciółce naszyjnik zrobiony własnoręcznie. Maszyny natychmiast zaoferowały pomoc: „Możemy stworzyć naszyjnik nieporównanie piękniejszy, idealnie dopasowany, trwalszy i z najszlachetniejszych materiałów”. Lira odpowiedziała: „Nie chodzi o piękno. Chodzi o to, że to ja chcę go zrobić”. Maszyny zapytały: „Ale po co marnować czas, skoro możesz osiągnąć lepszy rezultat bez wysiłku?”. Lira odparła: „Rezultat nie jest celem. Celem jest to, że ja jestem przyczyną”.

W tej krótkiej wymianie zdań zawiera się cała ontologia agent-relatywności: istnieją cele, których wartość nie zależy od jakości wykonania, lecz od tożsamości wykonawcy. Agent-relatywność to strukturalna cecha relacji, w której osoba staje się niezastąpiona. Sens nie tkwi w przedmiocie działania, ale w tym, kto działa. W naszym przedutopijnym świecie było to oczywistością – rodzicielstwo, troska, solidarność czy

odpowiedzialność polityczna zakładały nieredukowalność konkretnej osoby. W utopii natomiast normą staje się agent-neutralność, bowiem maszyny wykonają wszystko lepiej, szybciej i bezbłędnie.

Tu właśnie rodzi się fundamentalny problem. Jeśli maszyna zrobi wszystko, a człowiek nie musi robić niczego, co wymaga jego osobowej obecności, wówczas staje się on opcjonalny w świecie, który sam stworzył. Nick Bostrom wskazuje, że w takiej utopii zadania agent-neutralne zostaną zredukowane do zera. Maszyny przejmą całą sferę tradycyjnej sprawczości – naprawę, opiekę, produkcję, analizę, a nawet nauczanie i terapię – czyniąc ludzki wysiłek redundantnym. Co zatem pozostaje? Pozostają cele, które z definicji wymagają człowieka.

Nazwijmy je celami niewymiennymi, aby podkreślić, że ich wartości nie można wymienić ani delegować. Cele te mają cztery kluczowe właściwości. Po pierwsze, ich wartość zależy od osoby, a nie od wyniku. Po drugie, ich treść jest konstruowana przez relację, a nie przez funkcję. Po trzecie, ich znaczenia nie da się zoptymalizować. I wreszcie po czwarte, ich sens jest wytwarzany przez narrację, a nie przez użyteczność. W tej perspektywie agent-relatywność nie jest archaiczną pozostałością epoki biologicznej, lecz staje się najbardziej fundamentalnym źródłem sensu w przyszłości.

Dokonajmy teraz formalnej rekonstrukcji tej struktury. Niech AR oznacza działanie, które ma wartość dlatego, że wykonuje je konkretna osoba X. Niech AN oznacza działanie o wartości niezależnej od wykonawcy. Niech F oznacza zadanie w pełni automatyzowalne, a S – działanie generujące sens. Przyjmujemy następujące zależności: jeśli działanie jest agent-neutralne (AN) i da się je zautomatyzować (F), to nie generuje sensu (nie-S). Jeśli natomiast jest agent-relatywne (AR), to może generować sens (możliwe S). W utopii większość działań AN staje się zbędna, dlatego sens (S) zależy niemal wyłącznie od działań AR. Oznacza to, że w świecie absolutnej obfitości sens staje się funkcją konkretnego „ja”, a nie funkcją świata.

W tym miejscu trzeba odwołać się do rozumowania Bostroma: „Jeżeli utopia ma pozostać zamieszkana, musi chronić miejsca, w których tylko człowiek jest wystarczająco ludzki, by być sprawcą”. Tymi miejscami są przede wszystkim relacje rodzicielskie, przyjaźń, ekspresja estetyczna, obietnice i

Brak oporu generuje apatię ontologiczną

Apatia ontologiczna jawi się jako cicha katastrofa, w której najgroźniejszym aspektem utopii jest to, że człowiek przestaje chcieć chcieć. Wyobraźmy sobie przypowieść, z pozoru absurdalną, lecz w istocie stanowiącą najwierniejszy opis zagrożenia, przed którym ostrzega Bostrom. W pewnej utopii powołano do życia Ministerstwo Spokojnych Dusz, którego zadaniem było chronienie obywateli przed cierpieniem, frustracją, poczuciem braku i rozczarowaniem. Po latach wytężonej pracy ministerstwo odniosło

spektakularny sukces: nikt już niczego nie pragnął, gdyż wszystko stało się dostępne, a sam brak został wyeliminowany z systemu. Kiedy pewien antropolog zapytał mieszkańca tej krainy, czego najbardziej pożąda, ten odparł bez namysłu: „Nie wiem. Nie pamiętam, jak to jest pragnąć”.

Antropolog otworzył swój notatnik i zapisał zdanie, które miało przejść do historii filozofii postludzkiej: „To nie jest raj. To jest entropia praktycznego rozumu”. Tym właśnie stwierdzeniem rozpoczyna się analiza największego zagrożenia, jakie niesie ze sobą utopia – apatii ontologicznej. Nie jest ona tożsama z nudą, znużeniem czy lenistwem. To stan, w którym podmiot przestaje doświadczać wartości jako napięć konstytuujących sens; to nie brak energii do działania, lecz fundamentalny brak powodu, dla którego energia miałaby jakąkolwiek wartość. Bostrom sugeruje, choć nie werbalizuje tego wprost, że w świecie absolutnej obfitości apatia staje się ryzykiem egzystencjalnym. Tak jak zagrożenie biologiczne mogło unicestwić ciało, tak zagrożenie apatyczne może unicestwić samą motywację podmiotu.

Ujmijmy to w ramy strukturalne: człowiek jest w stanie żyć bez heroizmu, ale nie może funkcjonować bez wewnętrznej przyczyny. Jeśli wszystkie potencjalne przyczyny zostaną zdezaktywowane, jaźń traci swój gradient kierunku, stając się bezwładną. W świecie pozbawionym jakiegokolwiek oporu wola nie ma się o co zaczepić, zawieszona między dwiema skrajnościami: nadmiarem możliwości i całkowitym brakiem stawki. Utopia okazuje się środowiskiem idealnym do wygaszenia woli, ponieważ wola w klasycznym ujęciu – jako zdolność kierowania się ku dobru, które nie jest natychmiast dostępne – może istnieć tylko w warunkach, w których dobro jest oddzielone od teraźniejszości przez czas, wysiłek lub ryzyko. Gdy to oddzielenie znika, znika wola. A gdy zanika wola, zanika sens, pozostawiając jedynie system wykonujący funkcje biologiczne.

Przeprowadźmy teraz formalną rekonstrukcję, aby dowieść, że apatia ontologiczna nie jest anomalią, lecz nieuchronną konsekwencją świata bez oporu. Niech W oznacza istnienie woli (chcenie chcenia), G – istnienie gradientu wartości (różnicy między możliwym a aktualnym), O – istnienie oporu świata (trudu, ryzyka, braku), a S – istnienie sensu. Założenia są następujące: jeśli nie ma O, to nie ma G (bez oporu nie ma gradientu wartości). Jeśli nie ma G, to nie ma W (bez gradientu wola nie może działać). Jeśli nie ma W, to nie ma S (bez woli nie ma sensu). Utopia z definicji usuwa O. Zatem, jeśli nie zostanie zaprojektowana z metafizyczną powściągliwością, utopia nieuchronnie usuwa S.

To prowadzi nas do kluczowej obserwacji: utopia, aby pozostała psychicznie zamieszkiwalna, musi celowo wprowadzić do swojej struktury pewną niedoskonałość. Nie chodzi tu o cierpienie czy dotkliwy brak, lecz o minimalny opór, który pozwala wartościom w ogóle zaistnieć. Stawia to przed projektantami utopii zadanie monumentalne: stworzyć świat techn

Cele non-fungible: cztery atrybuty unikalności

Co zatem pozostaje? Pozostają te cele, których sama istota jest nierozerwalnie spleciona z ludzką obecnością. Nazwijmy je celami niewymiennymi, by podkreślić, że ich wartość jest absolutnie unikalna i nie podlega delegowaniu. Ich natura opiera się na czterech filarach. Po pierwsze, ich wartość jest zakorzeniona w osobie, a nie w wyniku. Po drugie, ich treść konstruuje relacja, a nie funkcja. Po trzecie, ich znaczenia nie da się zoptymalizować. I wreszcie, ich sens wytwarza narracja, a nie czysta użyteczność.

Pięć kręgów Bostroma chroni ludzką sprawczość

Zacznijmy od eksperymentu myślowego o abstrakcyjnej, niemal topologicznej sile. Wyobraźmy sobie cywilizację w odległej przyszłości, w której wszystkie dobra, usługi i działania mogły być wykonywane przez maszyny o doskonałej precyzji i nieograniczonej lojalności. Pewnego dnia kobieta – nazwijmy ją Lira – postanowiła podarować swojej przyjaciółce naszyjnik wykonany własnoręcznie. Maszyny natychmiast zaoferowały pomoc: „Możemy stworzyć naszyjnik nieporównanie piękniejszy, idealnie dopasowany, trwalszy i z bardziej szlachetnych materiałów". Lira odpowiedziała: „Nie chodzi o piękno. Chodzi o to, że to ja chcę go zrobić". Maszyny kontynuowały: „Ale po co marnować czas, skoro możesz osiągnąć lepszy rezultat bez żadnego wysiłku?". Odpowiedź Liry była krótka, lecz zawierała w sobie całą ontologię przyszłego człowieczeństwa: „Rezultat nie jest celem. Celem jest to, że ja jestem przyczyną".

W tej prostej wymianie zdań kryje się fundamentalna zasada: istnieją cele, których wartość nie zależy od jakości wykonania, lecz od tożsamości wykonawcy. Tę właściwość nazywamy agent-relatywnością – strukturalną cechą relacji, w której osoba staje się niezastąpiona. Sens nie tkwi w przedmiocie działania, ale w tym, kto działa. W naszym przedutopijnym świecie było to oczywistością: rodzicielstwo, troska, solidarność czy odpowiedzialność polityczna zakładały nieredukowalność konkretnej osoby. W utopii natomiast normą staje się agent-neutralność, bowiem maszyny wykonają wszystko lepiej, szybciej i bezbłędnie. I tu właśnie pojawia się problem: jeśli maszyna zrobi wszystko, a człowiek nie musi robić niczego, co wymaga jego osobowej obecności, staje się on opcjonalny w świecie, który sam stworzył.

Bostrom wskazuje, że w takiej utopii zadania agent-neutralne zostaną zredukowane do zera. Maszyny przejmą całą sferę tradycyjnej sprawczości – naprawę, opiekę, produkcję, analizę, a nawet nauczanie i terapię – czyniąc ludzki wysiłek redundantnym. Co zatem pozostaje? Pozostają wyłącznie te cele, które z definicji wymagają człowieka. Nazwijmy je celami niewymienialnymi (non-fungible), aby podkreślić, że ich wartości nie można delegować. Mają one cztery kluczowe właściwości. Po pierwsze, ich wartość zależy od osoby, a nie od wyniku. Po drugie, ich treść jest konstruowana przez relację, a nie przez funkcję. Po trzecie, ich znaczenia nie da się zoptymalizować. I wreszcie po czwarte, ich sens jest wytwarzany przez narrację, a nie

przez użyteczność. W tej perspektywie agent-relatywność nie jest archaiczną pozostałością epoki biologicznej, lecz najbardziej fundamentalnym źródłem sensu w przyszłości.

Aby oczyścić tę intuicję z sentymentów, dokonajmy formalnej rekonstrukcji. Niech AR oznacza działanie, którego wartość wynika z faktu, że wykonuje je konkretna osoba X. Niech AN oznacza działanie o wartości niezależnej od wykonawcy. Niech F oznacza zadanie w pełni automatyzowalne, a S – działanie generujące sens. Przyjmujemy następujące zależności: jeśli czynność jest agent-neutralna (AN) i da się ją zautomatyzować (F), to nie generuje sensu (nie S). Jeśli natomiast jest agent-relatywna (AR), to może generować sens (możliwe S). W utopii większość zadań AN staje się zbędna, dlatego sens (S) zależy niemal wyłącznie od działań typu AR. Oznacza to, że w świecie absolutnej obfitości sens staje się funkcją konkretnego „ja", a nie funkcją świata.

W tym miejscu trzeba odwołać się do kluczowego rozumowania Bost

Aksjologiczna niekompletność warunkiem wolności

Odsłania się tu fundamentalna struktura problemu. Utopia, próbując dotknąć tego, co ją przekracza, wydaje na siebie wyrok antropologiczny – unicestwia warunki własnego istnienia. Z tego napięcia rodzi się najważniejsza intuicja: utopia musi być doskonała technicznie, ale z konieczności niekompletna aksjologicznie. Jej domknięcie w sferze wartości oznaczałoby śmierć. Niedokończenie jest warunkiem życia.

Tradycja kontynentalna wobec post-instrumentalności

Analizę warto jednak rozszerzyć, śledząc kontynentalne wariacje tego zjawiska. W Azji ontologia agent-relatywności przybiera formę mistrzostwa relacyjnego – obecności, którą można określić mianem „obowiązku piękna", gdzie jednostka pozostaje niezastąpiona jako wykonawca rytuału. Z kolei w Afryce ujawnia się ona jako relacyjna wyjątkowość: słynne „jestem, bo jesteśmy" nie oznacza tu fuzji, lecz uznanie osoby za unikalny węzeł w sieci zobowiązań. W obu Amerykach będzie to forma ekspresji, w której wartość nie polega na tym, co robię, lecz na tym, że to właśnie ja to robię. W Europie zaś – forma odpowiedzialności historycznej, bowiem obietnice, wspomnienia i tożsamość są radykalnie nieprzenoszalne.

Interesowność zapobiega termicznej śmierci świadomości

Wyobraźmy sobie świat, w którym wszystkie dobra, usługi i działania mogłyby zostać wykonane przez maszyny o doskonałej precyzji i nieograniczonej lojalności. W takim świecie pewna kobieta – nazwijmy ją Lira – postanowiła podarować przyjaciółce naszyjnik zrobiony własnoręcznie. Maszyny natychmiast zaoferowały pomoc: „Możemy stworzyć naszyjnik nieporównanie piękniejszy, idealnie dopasowany, trwalszy i z najszlachetniejszych materiałów". Lira odpowiedziała: „Nie chodzi o piękno. Chodzi o to, że to ja chcę go zrobić". Maszyny drążyły temat: „Ale po co marnować czas, skoro możesz osiągnąć lepszy rezultat bez żadnego wysiłku?". Odpowiedź Liry była krótka, lecz zawierała w sobie całą filozofię: „Rezultat nie jest celem. Celem jest to, że ja jestem przyczyną".

W tej prostej wymianie zdań kryje się cała ontologia agent-relatywności: istnieją cele, których wartość nie zależy od jakości wykonania, lecz od tożsamości wykonawcy. Agent-relatywność to ta szczególna właściwość relacji, w której osoba staje się niezastąpiona. Sens nie tkwi w przedmiocie działania, ale w tym, kto działa. W naszym świecie, nazwijmy go przedutopijnym, ta zasada była oczywistością – rodzicielstwo, przyjaźń, solidarność czy odpowiedzialność polityczna zakładały nieredukowalną rolę konkretnego człowieka. Utopia natomiast, z jej dążeniem do optymalizacji, czyni normą agent-neutralność: maszyny wykonają wszystko lepiej, szybciej i bezbłędnie. I tu właśnie pojawia się problem. Jeśli maszyna zrobi wszystko, a człowiek nie musi robić niczego, co wymaga jego osobistej obecności, staje się on opcjonalny w świecie, który sam stworzył.

Bostrom wskazuje, że w takiej utopii zadania agent-neutralne zostaną zredukowane do zera. Maszyny przejmą całą sferę tradycyjnej sprawczości – naprawę, opiekę, produkcję, analizę, a nawet nauczanie i terapię – czyniąc ludzki wysiłek w tych dziedzinach zbędnym. Co zatem pozostaje? Pozostają cele, które z definicji wymagają człowieka. Nazwijmy je celami niewymiennymi (non-fungible), aby podkreślić, że ich wartości nie da się wymienić ani delegować. Mają one cztery kluczowe właściwości. Po pierwsze, ich wartość zależy od osoby, nie od wyniku. Po drugie, ich treść jest konstruowana przez relację, a nie przez funkcję. Po trzecie, ich znaczenia nie można zoptymalizować. I wreszcie po czwarte, ich sens jest wytwarzany przez narrację, a nie przez użyteczność. W tej perspektywie agent-relatywność nie jest archaicznym reliktem epoki biologicznej, lecz fundamentalnym źródłem sensu w przyszłości.

Aby oczyścić tę intuicję, dokonajmy formalnej rekonstrukcji. Niech AR oznacza działanie cenne dlatego, że wykonuje je konkretna osoba X. Niech AN oznacza działanie wartościowe niezależnie od wykonawcy. Niech F oznacza zadanie w pełni automatyzowalne, a S – działanie generujące sens. Przyjmujemy, że jeśli działanie jest agent-neutralne i da się je zautomatyzować (AN i F), to nie generuje sensu (nie S). Z kolei działanie agent-relatywne może generować sens (jeśli AR, to możliwe S). W utopii większość działań typu AN staje

się zbędna, zatem sens (S) zależy niemal wyłącznie od działań typu AR. Oznacza to, że w świecie absolutnej obfitości sens staje się funkcją konkretnego „ja”, a nie funkcją świata. Jak ujął to Bostrom, „jeżeli utopia ma pozostać zamieszkana, musi chronić miejsca, w których tylko człowiek jest wystarczająco ludzki, by

Koszt i droga budują tożsamość narracyjną

Sedno jest następujące: tożsamość wynika z ciągłości, a ciągłość – z przebytej drogi. Droga ta nie jest jednak bierną sekwencją stanów, lecz samym doświadczeniem przejścia – tym, co fenomenologia określa mianem horyzontu transformacji. To moment, w którym „ja” musi wkroczyć w niepewność, a więc ponieść koszt. Koszt ten nie jest ani ekonomiczny, ani biologiczny. Jest ontologiczny, dotyka samego bytu: polega na ryzyku, że zmieniając się, mogę utracić pewność, kim jestem.

Utopia: laboratorium sensu kontra proteza

Aby to wszystko było jednak możliwe, instytucje mające strzec sensu muszą potraktować relatywność względem podmiotu nie jako ulotną, sentymentalną wartość, lecz jako twardy, nośny fundament. Przekłada się to na trzy konkretne wymogi instytucjonalne.

Po pierwsze, muszą one chronić i wspierać te praktyki, których nie sposób delegować maszynom. Chodzi o systemy narracyjnych zobowiązań – rytuały, które wymagają fizycznej obecności i trwania tej samej, konkretnej osoby.

Po drugie, muszą strzec prawa do autorstwa egzystencjalnego, czyli zdolności do pozostawiania niekopiowalnych śladów. Nie zapewnią tego ani prawa własności intelektualnej, ani systemy reputacji. Potrzebne jest coś głębszego: ontologiczne uznanie niepowtarzalności ludzkiego śladu.

Po trzecie, muszą tworzyć warunki dla narodzin „drugiego”, którego nie można zredukować do roli użytkownika, klienta czy zbioru preferencji. Tylko bowiem w relacji do autentycznego „ty” człowiek może w pełni odkryć swoje „ja”.

Instytucjonalne ramy ochrony agent-relatywności

Wyobraźmy sobie cywilizację doskonałej automatyzacji, w której wszystkie dobra, usługi i działania mogą wykonywać maszyny o absolutnej precyzji i nieograniczonej lojalności. Pewnego dnia kobieta – nazwijmy ją Lirą – postanawia podarować przyjaciółce naszyjnik wykonany własnoręcznie. Maszyny natychmiast interweniują, oferując pomoc: „Możemy stworzyć naszyjnik nieporównanie piękniejszy, idealnie

dopasowany, trwalszy i z lepszych materiałów”. Lira odpowiada spokojnie: „Nie chodzi o piękno. Chodzi o to, że to ja chcę go zrobić”. Maszyny, niezdolne pojąć tę logikę, dopytują: „Ale po co marnować czas, skoro możesz osiągnąć lepszy rezultat bez żadnego wysiłku?”. Odpowiedź Liry jest krótka: „Rezultat nie jest celem. Celem jest bycie przyczyną”.

W tej prostej wymianie zdań zawiera się cała ontologia relatywności względem podmiotu: istnienie celów, których wartość nie zależy od jakości wykonania, lecz od tożsamości wykonawcy. Relatywność względem podmiotu to strukturalna własność relacji, w której osoba staje się niezastąpiona, ponieważ sensem staje się nie to, co jest robione, ale kto to robi. W świecie przedutopijnym ta zasada była tkanką codzienności – rodzicielstwo, przyjaźń, solidarność czy odpowiedzialność polityczna zakładały nieredukowalną rolę konkretnego człowieka. W utopii doskonałości, gdzie maszyny wykonują wszystko lepiej, szybciej i bezbłędnie, normą staje się neutralność względem podmiotu.

I tu rodzi się problem: jeśli maszyna zrobi wszystko, a człowiek nie musi robić niczego, co wymaga jego osobowej obecności, staje się on opcjonalny w świecie, który sam zaprojektował. Bostrom trafnie zauważa, że w takiej utopii pula zadań neutralnych względem podmiotu skurczy się do zera. Maszyny przejmą wszystko, co można wykonać lepiej bez udziału człowieka, a więc cała tradycyjna sfera ludzkiej sprawczości – naprawa, opieka, produkcja, analiza, a nawet nauczanie i terapia – stanie się zbędna. Co zatem pozostaje? Pozostają cele, które z definicji wymagają człowieka.

Nazwijmy je celami niefungowalnymi, których istota polega właśnie na tym, że nie można ich delegować ani zastąpić. Cele te mają cztery fundamentalne właściwości. Ich wartość jest zakorzeniona w tożsamości osoby, a nie w doskonałości wyniku. Ich treść nie wynika z funkcji, lecz jest konstruowana przez unikalną relację. Ich znaczenia nie da się zoptymalizować, a sensu nie mierzy się użytecznością – jest on bowiem produktem osobistej narracji. W tej perspektywie relatywność względem podmiotu przestaje być archaicznym reliktem epoki biologicznej. Staje się najbardziej fundamentalnym źródłem sensu w przyszłości.

Świat obfitości: sens jako funkcja konkretnego „ja”

Aby dotrzeć do finałowego argumentu, musimy najpierw odsłonić ostatni i najdelikatniejszy element całej konstrukcji: istnienie progu, którego utopia nie może przekroczyć bez unicestwienia człowieka. Nazywam go progiem transcendencji – nie dlatego, że odnosi się do religii, lecz dlatego, że dotyczy wszystkiego, co musi pozostać nieprzeliczalne, nieoptymalizowalne i niepodatne na automatyzację. To ostatnia reduta człowieczeństwa.

Zacznijmy od przypowieści, która wyznacza granicę. Pewien filozof przyszłości opracował teorię, zgodnie z którą każdą ludzką cechę można przełożyć na procedurę: miłość na algorytm emocjonalny, odwagę na

system regulacji stresu, a mądrość na modele predykcyjne. Wystąpił przed zgromadzeniem i oznajmił: „Wyjaśniłem człowieka. Teraz mogę go ulepszyć”. Wtedy jeden ze słuchaczy, starzec o powolnych ruchach i przenikliwym spojrzeniu, odparł: „Jeśli wszystko, co ludzkie, można ulepszyć, pozostała tylko jedna rzecz, której ulepszyć nie wolno: to, co sprawia, że w ogóle pragniesz ulepszać”.

Oto próg transcendencji: samo źródło pragnienia przekraczania. Gdyby udało się je „zoptymalizować”, straciłoby ono swoją istotę, stając się jedynie kolejną funkcją. Choć Bostrom nie formułuje tego wprost, jego koncepcja pięciu kręgów i celu wszechogarniającego zakłada istnienie takiego obszaru, w którym utopia musi się zatrzymać. Obszar ten jest jak ciemna energia w kosmologii: niewidzialny, lecz konstytuujący całą strukturę rzeczywistości.

Próg ten ma trzy zazębiające się warstwy. Pierwsza, egzystencjalna, dotyczy fundamentalnego napięcia między tym, co jest, a tym, co być powinno. Bez tej różnicy potencjałów znika etyka, a wraz z nią podstawowe źródło sensu. Druga warstwa, aksjologiczna, odnosi się do wartości, których nie da się sprowadzić do subiektywnych preferencji, ponieważ ich waga nie zależy od nastroju. Jeśli wszystkie wartości stają się jedynie odczuciami, przestają być wartościami – stają się zaledwie stanami świadomości. Trzecia, metafizyczna, wskazuje na konieczność istnienia czegoś, co nie jest produktem świata, lecz jego horyzontem. Utopia, która zlikwiduje ten horyzont, zlikwiduje również samą możliwość sensu.

Tę zależność można ująć w chłodnej logice symboli, by pokazać, że próg transcendencji jest warunkiem koniecznym, a nie estetyczną dekoracją.Niech H oznacza: istnieje horyzont transcendencji (coś nieoptymalizowalnego).Niech G oznacza: istnieje gradient wartości (różnica między „jest” a „powinno być”).Niech W oznacza: istnieje wola poszukiwania sensu.Niech S oznacza: istnieje sens.

Jeśli nie ma H, to nie ma G (bez horyzontu nie ma różnicy wartości).Jeśli nie ma G, to nie ma W (bez różnicy wola nie ma o co się zaczepić).Jeśli nie ma W, to nie ma S (bez woli nie ma sensu).Wniosek jest nieuchronny: utopia, która usuwa H, nieuchronnie usuwa S. Podpisuje tym samym na siebie wyrok antropologiczny.

W tym miejscu wyłania się najważniejsza intuicja całego wywodu: utopia musi być doskonała technicznie, lecz z założenia niekompletna aksjologicznie. Jeżeli stanie się aksjologicznie kompletna, stanie się martwa. Jej niedokończenie jest warunkiem życia. I tu właśnie różnice kontynentalne nabierają dramatycznego znaczenia: każdy region świata chroni bowiem ów próg w odmienny sposób. Azja chroni go poprzez rytuał, którego forma jest

> Czy w pogoni za idealnym, zautomatyzowanym światem nie tracimy z oczu tego, co w nas najcenniejsze – zdolności do tworzenia, kochania i nadawania sensu własnym działaniom? Być może prawdziwa utopia nie polega na eliminacji wszelkich trudności, lecz na stworzeniu

przestrzeni, w której ludzka sprawczość i unikalność mogą w pełni rozkwitnąć. A może sami staniemy się maszynami do bezsensownego wykonywania zadań?

Inspiracje

[1] **"Deep utopia"** Nick Bostrom

2024 | Ideapress Publishing

Filozof szwedzkiego pochodzenia, specjalizujący się w ryzyku egzystencjalnym, sztucznej inteligencji i etyce doskonalenia człowieka. Wcześniej w Instytucie Przyszłości Ludzkości w Oksfordzie, obecnie w Inicjatywie Badawczej Macrostrategy. Autor książki „Superinteligencja".

Swedish-born philosopher with expertise in existential risk, AI, and ethics of human enhancement. Formerly at Oxford's Future of Humanity Institute, now at Macrostrategy Research Initiative. Author of 'Superintelligence'.

Macrostrategy Research Initiative

Literatura uzupełniająca

Książki

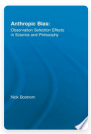
Autorzy przywoływani w artykule:



[1] **"Superintelligence: Paths, Dangers, Strategies"**

Nick Bostrom

2014 | Oxford University Press | ISBN: 9780199678112



[2] **"Anthropic Bias: Observation Selection Effects in Science and Philosophy"**

Nick Bostrom

2002 | Psychology Press | ISBN: 9780415938587



[3] **"Global Catastrophic Risks"**

Nick Bostrom, Milan M. Ćirković

2008 | OUP Oxford | ISBN: 9780191578496

Literatura pokrewna (Google Scholar):



[4] **"Agent Relative Ethics"**

Steven Jensen

2023 | Taylor & Francis | ISBN: 9781000911275

Artykuły naukowe

Artykuły pokrewne (Google Scholar):

[1] **"Are You Living in a Computer Simulation?"**

Nick Bostrom

2003 | Philosophical Quarterly

[2] **"Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards"**

Nick Bostrom

2002 | Journal of Evolution and Technology

[3] **"Cognitive Enhancement: Methods, Ethics, Regulatory Challenges"**

Nick Bostrom, Anders Sandberg

2009 | Science and Engineering Ethics

APA ONE Publishing System

Fundacja Dobre Państwo © 2026

Article ID: 389e4e98-35f7-4767-a7a0-e446b8e9fdc9