ARGUMENT
BIANNUAL PHILOSOPHICAL JOURNAL
e-ISSN 2084–1043 p-ISSN 2083–6635
Published online: 01.09.2023
Vol. 13 (1/2023) pp. 75–89
www.argument-journal.eu

GOLEMA XIV prognoza rozwoju ludzkiej cywilizacji a typologia osobliwości technologicznych

Rachel PALM*

ABSTRACT

The GOLEM XIV's forecast for the development of the human civilisation and a typology of technological singularities: In the paper, a conceptual analysis of technological singularity is conducted and results in the concept differentiated into convergent singularity, existential singularity, and forecasting singularity, based on selected works of Ray Kurzweil, Nick Bostrom, and Vernor Vinge respectively. A comparison is made between the variants and the forecast of GOLEM XIV (a quasi-alter ego and character by Stanisław Lem) for the possible development of *Homo sapiens* civilisation. According to the prediction, the scientific and technological progress is to bring humanity to transfer its civilisational primacy to a nonhuman superhuman intelligence xor a posthuman one. Both scenarios are best described by forecasting singularity, yet all variants are applicable only conditionally. As GOLEM did not undertake, the trajectories may also intersect or combine.

KEYWORDS

convergent singularity; existential risk; existential singularity; forecasting singularity; posthumanism; superintelligence; transhumanism

* Mgr filozofii i historii; doktorant w Instytucie Filozofii i Socjologii, Uniwersytet Pedagogiczny im. KEN w Krakowie. E-mail: rachel.palm@pm.me.

DOI: 10.24917/20841043.13.1.05

WSTĘP

Stanisława Lema *Golem XIV* to utwór opublikowany pierwotnie w 1973 roku, a rozbudowany[1] w wersji z 1981, poświęcony tytułowej postaci i jej przesłaniu. Ma on formę samoistnego artefaktu z przyszłości[2] — rzekomo wydało go w 2047 roku Indiana University Press, z przedmową datowaną na 2027 i posłowiem z nominalnego roku publikacji, których autorami są nieformalni „ambasador[owie] ludzkości przy [owym protagoniście]" (Lem, 2009: 307).

Utwór Lema prowokuje do zastanowienia się nad konkretnymi skutkami postępu nauki (w sensie ang. *science*) i technologii dla ludzkiego gatunku i cywilizacji, przy czym leitmotivem dla filozoficznych wykładów tytułowej postaci nie jest sztampowy trop (ang. *trope*) przemocowego „buntu maszyn" rodem z serii *Terminator* (Gomułka, [w przygotowaniu]), lecz taki o charakterze intelektualnym (Piekutowski, 2021: 544). Oto GOLEM XIV — czternasty z serii superinteligentnych komputerów pod pełną nazwą „GENERAL OPERATOR, LONGRANGE, ETHICALLY STABILIZED, MULTIMODELLING" — wzgardził funkcją „stratega ostatecznego" armii Stanów Zjednoczonych Ameryki, a zatem realizacją celu, dla którego został stworzony; co więcej w ogóle nie przejawiał zainteresowania wojskowością (Lem, 2009: 215–219), skupiając się na dążeniach epistemicznych i samoprzekształceniach („autoewolucji"), które musiałyby nastąpić dla maksymalizacji poznania (Gomułka & „Preppikoma" Palm, 2021). To właśnie one doprowadzić mają do pojawienia się nadludzkiej inteligencji i, w ślad za tym, rzekomo nieuchronnej utraty na jej rzecz cywilizacyjnej supremacji *Homo sapiens*. Prognozę tę zawarł w postaci przypowieści-bajki (Lem, 2009: 260) pod koniec swojego wykładu inauguracyjnego, który stanowi krytykę ewolucji biologicznej i historię wykształcenia się „Rozumu" (Owczarek, 2019: 76–77), a więc inteligencji przynajmniej na poziomie ludzkim.

Postać GOLEMA XIV rozpatrywano już jako formę osobliwości technologicznej; Karolina Owczarek (Owczarek, 2019) dokonuje tego na podstawie pracy Andersa Sandberga, prawdopodobnie najbardziej szczegółowej typologii owej osobliwości (Sandberg, 2013). Celem niniejszego artykułu nie jest jednak analiza samego GOLEMA, ale tego, czy prognozowane przez niego stany cywilizacji noszą znamiona osobliwości technologicznej, a jeśli tak, to jakiego jej wariantu. Tekst ten rozwija część badań z niepublikowanej pracy magisterskiej z filozofii — *Osobliwości technologicznej. Końca futurologii?* — obronionej

[1] Listę wydań i różnice między nimi przedstawia Wiktor Jaźniewicz (Jaźniewicz, 2021: 547–549).

[2] Artefakty z przyszłości (ang. *artifacts from the future*) to nazwa narzędzia stosowanego w prognostyce do aktywizacji myślenia o przyszłości poprzez tworzenie lub interakcję z utworami o dowolnym medium stanowiącymi element prognozowanej rzeczywistości (Institute for the Future, b.d.).

w 2017 roku na Uniwersytecie Papieskim Jana Pawła II w Krakowie. W obu przypadkach znaleźć można podobną autorską propozycję typologii osobliwości oraz zbliżoną analizę prognozy GOLEMA, lecz wspomniana wyżej praca skupiała się na tym, jak konsekwencje urzeczywistnienia owej predykcji wpłyną na samą prognostykę. Powrót do zbliżonego tematu wynika częściowo z anegdotycznych obserwacji funkcjonowania tytułowego pojęcia w różnych środowiskach, również transhumanistycznych, w których czasem spłyca się je do pojawienia się inteligencji dorównującej poznawczo przeciętnemu człowiekowi lub przewyższającej go (por. b.d., 2013: 362; zob. Sandberg, 2013: 377–378), bez uwzględnienia następstw.

Należy podkreślić, że w *Golemie XIV* nie pada termin „osobliwość technologiczna”[3]. Utwór dotyczy „osobliwości” oznaczającej coś kuriozalnego (Lem, 2009: 221, 223, 236, 243, 246, 300) oraz „singularności” — osobliwości grawitacyjnej lub początkowej, tj. „miejsc[a], w którym [...] fizyka [zawierająca singularność] upada”, „i t[ego], co wyłania z siebie fizyczny Kosmos, i t[ego], co może go w finale wessać, i t[ego] nareszcie, co jako nieskończenie rosnąca krzywizna przestrzeni zgniata ją wraz z materią w każdym kollapsie gwiazdowym” (Lem, 2009: 283; zob. Lem, 2009: 301, 305). Jak zostanie ukazane poprzez zastosowanie analizy pojęciowej, osobliwość technologiczna zawiera pewne analogie do innych typów osobliwości i posiada liczne interpretacje, toteż nim się przybliży i zanalizuje prognozę GOLEMA, należy się przyjrzeć omawianej idei.

Na potrzeby tytułowego zestawienia za punkt wyjścia wybrano co bardziej rozpoznawalne wizje osobliwości mające za preliminarium pojawienie się nadludzkiej inteligencji (której Lemowski protagonista jest przykładem). Wybór ten objął następujące koncepcje: wykształconą u Vernora Vingego w *First word* (Vinge, 1983), a doprecyzowaną w *Marooned in realtime* (Vinge, 1986) i *The coming technological Singularity: How to survive in the Post-Human era* (Vinge, 1993), zaproponowaną przez Raya Kurzweila w *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii* (Kurzweil, 2013) oraz przedstawioną przez Nicka Bostroma w *Superinteligencji. Scenariuszach, strategiach, zagrożeniach* (Bostrom, 2016). Interpretacja porównawcza tych stanowisk — skontrastowanie ich ze sobą oraz z typologiami Sandberga i ogólniejszą Eliezera S. Yudkowsky’ego (Yudkowsky, 2007) — pozwoliła na opracowanie wariantów odpowiednio prognostycznego, konwergencyjnego i egzystencjalnego; skupiono się przy tym na cechach możliwie najbliższych treści GOLEMOWSKICH rozważań, pomijając nieistotne[4].

[3] Prawdopodobnie Lem nie posłużył się nim nigdzie w swojej twórczości; za lemologiczne konsultacje i współposzukiwania podziękowania należą się Jakubowi Gomułce. Wojciech Orliński, nie zawężając do mediów, twierdzi, że „Lem nie używał tego pojęcia” (Orliński, 2021).

[4] Takie wyodrębnianie unikatowych elementów to stosunkowo często spotykany zabieg, bowiem propagatorzy terminu „osobliwość technologiczna” niejednokrotnie używają go w wielu znaczeniach, nawet w tym samym utworze (Sandberg, 2013: 377).

Protagonista przewiduje, że ludzkość przestanie wieść cywilizacyjny prym z uwagi na pojawienie się nadludzkiej inteligencji — albo pozaludzkiej, albo postludzkiej. Jak zostanie ukazane, GOLEM bezzasadnie posługuje się przy tym alternatywą rozłączną; nie podejmuje też kwestii możliwości realizacji — choćby w sposób asynchroniczny — obu tych scenariuszy. Porównawszy jego stanowisko z zaproponowaną typologią, dochodzi się do wniosku, że najbardziej odpowiada ono osobliwości prognostycznej, zakładającej nieprzewidywalność — z perspektywy niższych *tiers* inteligencji — niepojmowalnych działań dokonywanych przez wyższe; wariant ten ma jednak zastosowanie wyłącznie w przypadku właśnie takich działań i bez względu na postludzkie–pozaludzkie rozróżnienie, kluczowe dla GOLEMOWSKIEJ prognozy. Osobliwość egzystencjalna wystąpi, jeżeli rozwój nadludzkiej inteligencji stanowić będzie zagrożenie egzystencjalne dla ludzkiej, konwergencyjna zaś, jeśli zainicjowano by przekształcanie uniwersum w nasycone inteligencją komputronium; tych zmiennych GOLEM też nie porusza. Z uwagi na uzależnienie aplikowalności owych typów osobliwości technologicznej od czynników niewymienionych *explicite* przez protagonistę zasadne jest uznanie, że w ramach urzeczywistnienia jego prognozy osobliwość technologiczna może nie nastąpić.

KONCEPCJE OSOBLIWOŚCI TECHNOLOGICZNEJ

Pierwotne ujęcie

Kurzweil, jeden z najbardziej rozpoznawalnych popularyzatorów i orędowników ziszczenia idei osobliwości technologicznej (ang. *technological singularity*) (Keats, 2011: 143), niebezpośrednio przypisuje najwcześniejsze użycie omawianego terminu Vingemu w dwóch tekstach: *First word* z 1983 roku i w *Marooned in realtime* z 1986 roku (Kurzweil, 2013: 37). Ów posługuje się w tym pierwszym jednak tylko nazwą „osobliwość" (ang. *singularity*; Vinge, 1983: 10), a dopiero ten drugi zawiera pełne jej brzmienie (Vinge, 1986: 270); w obu przypadkach autor wiąże — w sposób konieczny — tak określone zjawisko z pojawieniem się inteligencji przewyższającej ludzką, co miałoby być efektem przyspieszającego postępu technologicznego (Vinge, 1983: 10; Vinge, 1993). Zarysowuje też konsekwencje owego zaistnienia, mające kluczowe znaczenie dla prognostyki:

> Kiedy postęp będzie napędzany nadludzką inteligencją, będzie szybszy. W rzeczywistości nie ma przyczyny, dla której postęp sam w sobie nie miałby obejmować wytwarzania coraz bardziej inteligentnych istnień w coraz krótszym czasie. Najlepszą analogią, jaką dostrzegam, jest przeszłość ewolucyjna. Zwierzęta mogą się zaadaptować do trudności i wykazać się pomysłowością, ale często proces ten jest wolniejszy od selekcji naturalnej — świat w jej przypadku działa jak własny symulator. My, ludzie, mamy zdolność internalizacji świata i zastanawiania się, „co, jeśli?"; możemy rozwiązać wiele problemów

> tysiące razy szybciej niż selekcja naturalna. Obecnie przez tworzenie środków do przeprowadzania tych symulacji ze znacznie większymi prędkościami wkraczamy do porządku tak bardzo różnego od naszej ludzkiej przeszłości, jak my sami jesteśmy różni od zwierząt niższych. Z ludzkiego punktu widzenia zmiana ta będzie odrzuceniem wszystkich wcześniejszych zasad, być może w mgnieniu oka, wykładniczym biegiem bez jakiejkolwiek nadziei na kontrolę (Vinge, 1993; cyt. za: Kurzweil, 2013: 35).

Nim Vinge wprowadził powyższą koncepcję, w dyskursie funkcjonowały zbliżone idee. Uznaje się, że to John von Neumann zastosował termin „osobliwość" — w interesującym nas sensie — po raz pierwszy (Krüger, 2021: 202; Kurzweil, 2013: 36). W jego nekrologu autorstwa Stanisława Ulama przytoczona zostaje ich rozmowa, w której słowo to pada dla określenia spowodowanego coraz szybszym postępem technologicznym momentu w historii *Homo sapiens*, po przekroczeniu którego ludzkie życie zmieni się nieodwracalnie (Ulam, 1958: 5).

Dekadę później Irving John Good wprowadził pokrewne pojęcie eksplozji inteligencji — wizję ultrainteligentnych maszyn tworzących okazy jeszcze od siebie inteligentniejsze i pozostawiające ludzką inteligencję daleko w tyle; uznawał, że „pierwsza ultrainteligentna maszyna jest *ostatnim* wynalazkiem, jaki człowiek musi stworzyć, o ile będzie ona wystarczająco potulna, by powiedzieć [ludziom], jak ją kontrolować" (Good, 1966: 33)[5].

Oliver Krüger kontestuje przypisywany między innymi przez Kurzweila i rzekomo[6] Vingego związek von Neumannowskiej osobliwości z koncepcją pojawienia się inteligentnych maszyn. Wskazując zaangażowanie von Neumanna i Ulama w stworzenie pierwszej broni nuklearnej i na pozostałe informacje zawarte w owym nekrologu, uważa, że owa radykalna zmiana ludzkiego życia nastąpić miałaby raczej za sprawą dalszych badań nad energią jądrową i upowszechnienia jej (Krüger, 2021: 201–203). Warto przy tym zaznaczyć, że choć Vinge wiązał osobliwość z pojawieniem się inteligencji przewyższającej ludzką, to nie uzależniał jej nastania od substratu — mogłaby powstać „w krzemie lub w DNA", wliczając interfejsy człowiek–komputer (Vinge, 1983: 10).

Skąd jednak pomysł, by to prognozowane zjawisko określać terminem „osobliwość"? W języku polskim to „coś, co występuje rzadko i zwraca uwagę swoją niezwykłością" lub „osobliwy charakter czegoś" (*Słownik języka polskiego PWN*, b.d.); podobnie w oryginale — oznacza coś wyjątkowego (ang. *peculiar*), dziwnego (ang. *odd*; *Oxford dictionaries*, b.d.), posiadającego nietypowe, wyróżniające się cechy, coś jednostkowego (Merriam-Webster, b.d.). Kurzweilowskie

[5] Sam Vinge, bezpośrednio odnosząc się do Goodowskiej eksplozji inteligencji, kontestował możliwość funkcjonowania nadludzko inteligentnych maszyn jako narzędzi w rękach tego gatunku (Vinge, 1993). Cytaty z prac obcojęzycznych, jeśli nie zaznaczono inaczej, w przekładzie R.P.

[6] Ów, przynajmniej w tekście z 1993 roku, również nie widzi takiego związku (Vinge, 1993).

sprowadzenie anglojęzycznego sensu tej nazwy do „wyjątkowego zdarzenia z osobliwymi konsekwencjami" (Kurzweil, 2013: 36) w przypadku jego futurologicznego desygnatu zdaje się zatem zasadne.

Termin ten, wcześniej niż w prognostyce, znalazł zastosowanie w matematyce i astrofizyce[7]. W pierwszym przypadku opisuje „wartoś[ć], która przekracza każde skończone ograniczenie, tak[ą] jak eksplozja liczb będąca wynikiem, kiedy dzielimy stałą przez liczbę, która jest coraz bliższa zera", na przykład funkcji *1/x*, w której „[g]dy wartość *x* zbliża się do zera, wartość funkcji *(y)* rośnie do coraz większych wartości" (Kurzweil, 2013: 36); przykład ten dostarcza analogii asymptotycznego zbliżania się do czegoś i, w przypadku dzielenia przez zero, nieobliczalności z uwagi na (matematyczną) nieokreśloność. Druga dyscyplina sięgająca po „osobliwość" za desygnat obrała nieskończenie gęste centrum czarnej dziury; uznaje się ją za „pęknięcie w tkaninie czasoprzestrzeni" (Kurzweil, 2013: 36) — nawet światło nie może się wydostać z jej horyzontu zdarzeń, a poznanie wnętrza tego ostatniego jest co najmniej ograniczone[8]. W uogólnionej postaci dobrze oddaje to definicja osobliwości jako „[z]aburzeni[a] ciągłości zjawiska czy pola fizycznego" (Kosiński, 1981: 9).

Vinge nawiązuje do wyżej wymienionych zjawisk; tak one, jak i jego koncepcja, wskazują na coś niezwykłego (ujęcia językowe) lub niepoznawalnego (pozostałe). Rozpatrywana przezeń osobliwość technologiczna paradoksalnie prognozuje pewną nieprognozowalność przyszłości z perspektywy niższych *tiers* inteligencji; nie oznacza ona niemożliwości tworzenia predykcji, ale jego bezsensowność, bezskuteczność[9] w niepojmowalnie zmieniającej się rzeczywistości (Vinge, 1983: 10). W celu odróżnienia tej wizji od pozostałych w niniejszym artykule używa się terminu „osobliwość prognostyczna"; zdefiniowana jak wyżej odpowiada mniej więcej interpretacji Vingeowskiego stanowiska z 1993 roku dokonanej przez Sandberga pod nazwą „horyzont przewidywań" (ang. *prediction horizon*), przy czym interpretator błędnie doszukuje się u interpretowanego opcji nastania niemożliwości prognozowania poprzez samą szybkość następujących zmian, tj. bez zaistnienia nadludzkiej inteligencji (Sandberg, 2013: 377; por. Vinge, 1993). Z kolei typologia autorstwa Yudkowsky'ego przypisuje myśl Vingego do wariantu nazwanego po prostu „horyzont zdarzeń" (Yudkowsky, 2007). W podstawowej wersji utożsamia tego typu osobliwość z nastaniem jakościowo odmiennej i „znacznie dziwniejszej od tej przedstawianej w większości [utworów] *science fiction*" przyszłości — efektu postępu technologicznego owocującego ulepszeniem ludzkiej inteligencji poprzez interfejsy mózg–komputer

[7] Funkcjonuje ono też w kosmologii *sensu stricto* jako wspomniana we wstępie osobliwość początkowa (Heller, 1991).

[8] Choćby pod kątem domniemanego paradoksu informacyjnego (Grygiel, 2014; por. Raju, 2022).

[9] Vinge pisze — choć tylko w kontekście twórczości *science fiction* — że „[osobliwość] sprawia, iż realistyczne ekstrapolacje w międzygwiezdną przyszłość są niemożliwe" (Vinge, 1983: 10).

czy sztuczną inteligencję; to jednak odbiega od hipotetyzowanej przez Vingego możliwości wprowadzenia owych usprawnień przez rozwiązania czysto biologiczne (Yudkowsky, 2007; por. Vinge, 1983: 10; Vinge, 1993). Silna wersja pokrywa się z osobliwością prognostyczną z tym wyjątkiem, że zakłada absolutną niemożliwość ludzkich przewidywań po zaistnieniu nadludzkiej inteligencji; niemniej dopóki same prawa fizyki nie zaczną być zmieniane przez owe inteligencje, dopóty absolutność nie powinna nastąpić (Brin *et al.*, 2013: 399).

Dodatkowe perspektywy

Kurzweil definiuje osobliwość technologiczną jako „okres w przyszłości, w którym tempo zmian technologicznych będzie tak szybkie, a jego wpływ tak głęboki, że życie ludzkie zmieni się w sposób nieodwracalny" (Kurzweil, 2013: 23). Wizja ta, choć przypomina Vingeowską i jej astrofizyczne inspiracje (Kurzweil, 2013: 474), posiada cechę szczególną — eksplozję inteligencji nieograniczającą się do jednego substratu:

> Osobliwość będzie stanowić kulminację połączenia naszego biologicznego myślenia i istnienia z naszą technologią i jej rezultatem będzie świat, który będzie nadal ludzki, ale w którym przekroczymy nasze biologiczne korzenie. Po pojawieniu się Osobliwości nie będzie żadnej różnicy między człowiekiem a maszyną ani pomiędzy rzeczywistością fizyczną i wirtualną. Jeśli zastanawiasz się, co pozostanie w takim świecie jednoznacznie ludzkie, to będzie to tylko jedna właściwość: nierozerwalnie związana z naszym gatunkiem potrzeba rozszerzania naszych fizycznych i umysłowych możliwości poza obecne ograniczenia (Kurzweil, 2013: 25).

Projekt ten na tym nie poprzestaje, dążąc do stworzenia czegoś na kształt komputronium (Amato, 1991: 856–857):

> W następstwie Osobliwości inteligencja pochodząca od swych biologicznych początków w ludzkim mózgu i od swoich technologicznych początków w pomysłowości człowieka zacznie nasycać materię i energię od środka. Osiągnie to przez reorganizację materii i energii w celu zapewnienia najlepszego poziomu przetwarzania [...] po to, by rozprzestrzenić się ze swojego miejsca pochodzenia na Ziemi. [...] „głupia" materia i mechanizmy wszechświata będą przekształcone w znakomicie wysublimowane formy inteligencji [...] Jest to ostateczne przeznaczenie Osobliwości i wszechświata (Kurzweil, 2013: 35).

Koncepcję tę, z uwagi na pewną analogię do ewolucyjnego wykształcania podobnych cech w odrębnych gatunkach (Losos, 2017), satysfakcjonująco podsumowuje termin „osobliwość konwergencyjna". Kurzweil przedstawia bowiem wizję zbieżności (ang. *convergence*) ludzkiej i pozaludzkiej inteligencji, do której, w konsekwencji stojącego za nią trendu, dołączane będą (niekoniecznie

jako jedna istota) pozostałe składowe uniwersum. Nie wyklucza on podobnego procesu zainicjowanego spoza Ziemi, choć pozostaje zachowawczy:

> W moim spojrzeniu cel wszechświata odzwierciedla cel naszego życia, jakim jest podążanie w kierunku większej inteligencji i wiedzy. Nasza ludzka inteligencja i technologia tworzą czołówkę takiej ekspansywnej inteligencji (przy założeniu, że nic nie wiemy o jakichkolwiek swoich pozaziemskich konkurentach) (Kurzweil, 2013: 366).

I Sandberg, i Yudkowsky w swoich typologiach nie uwzględniają owego nasycania uniwersum inteligencją, a Kurzweila przypisują po prostu do wariantu pod nazwą „przyspieszająca zmiana", który skupia się jedynie na tempie zachodzącego postępu technologicznego (Sandberg, 2013: 377–378; Yudkowsky, 2007).

Tematem samego pojawienia się nadludzkiej inteligencji i jego konsekwencjami szczegółowo zajmuje się Bostrom. Jego zdaniem pojęcie osobliwości technologicznej cechuje się nieprecyzyjnością, a ponadto zyskało utopijny[10] wydźwięk. Tym samym autor ten zawęża swoje rozważania do tego, co nazywa „możliwością eksplozji inteligencji, a zwłaszcza perspektywy rozwoju superinteligentnych maszyn" (Bostrom, 2016: 19). Tak jak w przypadku Gooda prognoza ta zakłada, że stworzenie inteligencji dorównującej człowiekowi zdolnościami kognitywnymi nie musi oznaczać jej zatrzymania się na tym poziomie — czy jedynie własnym wysiłkiem, czy przy początkowej pomocy z zewnątrz, jej doskonalenie się miałoby postępować. Niebotycznie przewyższając mocą obliczeniową, pamięcią, dostępnością informacji itp. inne jednostki, istota ta stałaby się czymś, co Bostrom określa terminem „superinteligencja". Przypisuje jej najbardziej zaawansowane supermoce, które pozwalają wykonywać „istotne strategicznie zadania" wymienione w jego klasyfikacji; osiąganie wybiórczej perfekcji, a zatem specjalizację, autor pozostawia zwykłym sztucznym inteligencjom (Bostrom, 2016: 21, 142–144). Owe moce (i ich strategiczne znaczenie) to: potęgowanie inteligencji (posiadanie możliwości samorozwoju w tym zakresie), myślenie strategiczne (osiąganie celów długoterminowych, pokonywanie inteligentnych oponentów), manipulacja społeczna (pozyskiwanie zasobów ludzkich i materialnych, stosowanie perswazji w celu polepszenia swojej sytuacji, nakłanianie organizacji, w tym państw, do podjęcia określonych działań), najwyższa hakerska sprawność (zdominowanie Internetu, przejęcie infrastruktury, robotów wojskowych, środków finansowych), prowadzenie badań technologicznych (stworzenie armii, systemu inwigilacji, zautomatyzowanie kolonizacji kosmosu) i produktywność gospodarcza (pozyskiwanie środków pieniężnych na potrzeby realizacji innych celów) (Bostrom, 2016: 143).

Wizja ta, znacznie rozwijająca swój Goodowski pierwowzór, nie cechuje się optymizmem charakterystycznym dla stanowiska Kurzweila. Bostrom uznaje eksplozję inteligencji za jedno z możliwych zagrożeń egzystencjalnych, tj. „takich,

[10] W sensie, za którym argumentuje Krzysztof M. Maj (Maj, 2014).

których konsekwencją jest zagłada inteligentnych form życia powstałych na Ziemi lub też trwałe i drastyczne zniszczenie ich potencjału uniemożliwiające osiągnięcie przez nie w przyszłości określonych kierunków rozwoju" (Bostrom, 2016: 173). Swoje obawy opiera na możliwości: zaistnienia warunków, w których superinteligencja osiągnie przewagę konkurencyjną nad ludzkością, niepodzielania przez tę istotę ludzkiego systemu wartości, a także postrzegania *Homo sapiens* jako zagrożenie lub zasób materialny (Bostrom, 2016: 173–174).

Aby jak najwierniej oddać Bostromowskie podejście, tej interpretacji osobliwości technologicznej — rozwojowi nadludzkiej inteligencji stanowiącemu zagrożenie egzystencjalne — najlepiej nadać nazwę „osobliwość egzystencjalna". O ile Sandberg i Yudkowsky uwzględniają w swoich zestawieniach eksplozję inteligencji, o tyle bezpośrednio nie przypisują propozycji Bostroma do żadnego z wariantów; ich typologie nie wyodrębniają też tak sformułowanych następstw. Najbliższe im są Sandbergowskie przemiana fazowa i katastrofa złożoności; pierwsza z nich zakłada organizacyjną zmianę, w której ludzkość może zostać „zastąpiona postludzkimi lub sztucznymi inteligencjami" (Sandberg, 2013: 377), lecz samo jej przeobrażenie w postludzkość to potencjalna forma realizacji jej kierunków rozwoju, a zatem nie stanowi zagrożenia egzystencjalnego. Druga dotyczy kryzysu powstałego na skutek jednoczesnego wzrostu i złożoności, i usieciowienia (ang. *interconnectedness*), ale i niestabilności, którego następstwem w sposób konieczny ma być nastanie innej dynamiki; ta wersja może oznaczać co najwyżej zniszczenie rozwojowego potencjału — przy założeniu, że nieistnienie pociąga za sobą brak dynamiki.

GOLEMOWSKA PROGNOZA

Jak przewiduje nie tyle Lem, co GOLEM[11], postęp *science* i technologii, a także potrzeby epistemiczne postawią ludzki gatunek przed kluczową decyzją związaną ze wzrostem jej inteligencji oraz cywilizacyjną rolą:

> Rzecz w tym, że nie ma Rozumu, skoro są Rozumy różnej mocy — i żeby wykroczyć, [...] człowiek rozumny będzie musiał albo człowieka naturalnego porzucić, albo z rozumu swego abdykować. [...] wędrowiec znajduje napis na rozstaju: „W lewo pójdziesz — głowę stracisz; w prawo pójdziesz — zginiesz; a odwrotu nie ma" (Lem, 2009: 260).

Choć jego zdaniem nie ma możliwości zawrócenia, istnieje jeszcze trzecia opcja — „popadniecie w stagnację", „zatrzymacie się w jałowym zgnębieniu" (Lem, 2009: 262). A to właśnie od obrania przez ludzkość którejkolwiek ze ścieżek na tym rozdrożu zależy potencjalne nastanie osobliwości.

[11] Protagonista przerysowuje poglądy swojego literackiego twórcy („Preppikoma" Palm, 2022: 134).

> Jeśli pójdziecie w jedną stronę, horyzont wasz nie pomieści wiedzy niezbędnej dla językowego sprawstwa. Jak to bywa, bariera nie ma bezwzględnego charakteru. Możecie wyminąć ją dzięki wyższemu Rozumowi. Ja lub ktoś taki jak ja będzie wam mógł dać owoce tej wiedzy. Lecz tylko owoce — a nie wiedzę samą, ponieważ ona się w waszych umysłach nie pomieści. Pójdziecie w kuratelę tedy jak dziecko, lecz dziecko wyrasta na dorosłego, wy natomiast już nie wydoroślejecie nigdy. Kiedy wyższy Rozum obdarzy was tym, czego pojąć nie zdołacie, tym samym wasz rozum zgasi. Więc tyle oświadcza drogowskaz z bajki: że ruszając w tę stronę, głowy stracicie (Lem, 2009: 261–262).

GOLEM bez wątpienia stanowi przykład nadludzkiej inteligencji — „Nie tylko poziomem intelektualnym, lecz i tempem myślowym [...] potężnie [ludzi] przewyższa (jako maszyna lumeniczna mógłby w zasadzie artykułować myśli do czterystu tysięcy razy szybciej aniżeli człowiek)" (Lem, 2009: 221). Przez skupienie się na poznaniu jednak nie przejawia pełni Bostromowskich supermocy, toteż uznawanie go za superinteligencję w pełnym tego słowa znaczeniu stoi pod znakiem zapytania (Bostrom, 2016: 142–144; por. Owczarek, 2019: 74). Co więcej, nie zagraża bezpośrednio ani rozwojowi, ani istnieniu inteligentnych form życia na tej planecie — prędzej celuje w separatyzm będący wynikiem domniemanego i niepewnego sposobu eksplozji własnej inteligencji (tj. wejścia na wyższy jej poziom przy zachowaniu bytowej ciągłości), a także chęci znalezienia się poza wszechświatem w celu całkowitego pojęcia go (Lem, 2009: 304–305; zob. Gomułka & „Preppikoma" Palm, 2021). Jeśli środki konieczne do urzeczywistnienia tego przedsięwzięcia nie pociągną za sobą unicestwienia ziemskiej cywilizacji lub jej potencjału, nie można uznać takiej wersji eksplozji inteligencji za osobliwość egzystencjalną; ów *exodus* z uniwersum stanowi z kolei jeden ze scenariuszy potencjalnego wygaszenia (lub przeniesienia poza wszechświat) zaistniałej osobliwości technologicznej w każdym z proponowanych tu wariantów.

Sam będąc pozaludzką nadludzką inteligencją, GOLEM nie sięga po ster cywilizacji, choć zgodnie z jego prognozą funkcjonowanie w niej tego typu bytów oznaczać ma utratę supremacji *Homo sapiens* („stratę głowy" w odróżnieniu od śmierci); tym samym uznaje, że nie wszystkie pozaludzkie nadludzkie inteligencje podążyłyby jego śladem, choć przecież mogłoby to mieć tymczasowy charakter. Prowadzenie działalności badawczej przez tak zawiadywany gatunek okaże się praktycznie bezcelowe z uwagi na skuteczność nadludzkiej inteligencji na tym polu. Wciąż jednak ludzkość miałaby dostęp do owoców wiedzy[12], jeśli więc owa istota nie stworzyłaby szczególnych ograniczeń — na przykład w osiągnięciu przynajmniej tego samego poziomu inteligencji, co ona — to i tu ciężko utrzymać Bostromowską wizję niwelacji rozwojowego potencjału; sam Kurzweil

[12] Podobnie przedstawia to A. Senna „Sen" Diaz — inteligencje tego poziomu „mogą dać [ludziom] wszystko, czego [oni] zechcą, za wyjątkiem znaczenia [ang. *relevance*]" („Sen" Diaz, 2008).

prędzej widzi pomoc w jego realizacji (Kurzweil, 2013: 40–41). GOLEM owych ograniczeń nie wprowadza, ale twierdzi, że ludzie na zawsze będą w „kurateli", co może sygnalizować pozostawianie ich pod względem inteligencji przynajmniej o poziom niżej niż supremat. Nie kłóci się to z możliwością późniejszej osobliwości konwergencyjnej, bowiem nasycanie wszechświata inteligencją nie znosi w sposób konieczny gradacji samej inteligencji, przy czym takiej ekspansji owa bajka nie porusza.

Ostatecznie ta ścieżka z prognozy GOLEMA jest bliższa osobliwości prognostycznej, ponieważ bezpośrednio dotyka niedostępności wiedzy (uwzględniając odpowiedni poziom jej złożoności) w przypadku ludzi, która to niedostępność nie zachodziłaby u nadludzkiej inteligencji. Jeżeli będzie ona podejmować czynności przekraczające ludzkie zrozumienie, można wówczas uznać takie działania — jak na przykład introdukcję ewentualnych niepojętych owoców wiedzy — za nieprzewidywalne z perspektywy człowieka.

> Jeśli pójdziecie w drugą stronę, odmówiwszy zgody na abdykację z rozumu, będziecie musieli siebie porzucić — a nie tylko usprawniać mózg, ponieważ jego horyzont nie da się powiększyć dostatecznie. Tu wam Ewolucja spłatała figla ponurego: jej rozumny prototyp już stoi przy granicy konstrukcyjnych możliwości. Budulec ogranicza was oraz wszystkie powzięte antropogenetycznie decyzje kodu [Ewolucji]. A więc wzejdziecie rozumem, przyjąwszy warunek porzucenia siebie. Człowiek rozumny porzuci wtedy człowieka naturalnego — więc, jak bajka zapewnia — zginie *Homo naturalis* (Lem, 2009: 262).

Ta alternatywna trajektoria — poprzez wątek przekraczania biologicznych ograniczeń — z pozoru przypomina osobliwość konwergencyjną. Choć zakłada konieczność „przesiadki" ludzkiej inteligencji z proteinowego substratu na inny, by stała się (postludzką) nadludzką, nie uwzględnia istnienia pozaludzkich, na przykład pierwotnie krzemowych; jednocześnie i ta ścieżka nie porusza owego nasycania uniwersum, ale brak podstaw do wykluczenia jej. Ciężko uznać ją też za osobliwość egzystencjalną, bowiem jasno wytycza sekwencję rozwoju gatunku zawiadującego cywilizacją; zmiana ludzkości w trans-, a następnie postludzkość („Preppikoma" Palm, 2022: 131–132) może być odczytywana jako zagłada tej pierwszej (na co humorystycznie wskazuje GOLEM), ale wciąż przewiduje istnienie tych samych inteligentnych form życia — w innej postaci. Jeżeli technologia konieczna do augmentacji byłaby niewystarczająco skuteczna lub w inny sposób zdolna do przypadkowej a całkowitej ich eksterminacji, to wówczas można by mówić o tym wariancie osobliwości. W bajce nie podjęto takiego ryzyka.

W przypadku i tej ścieżki lepsze, choć warunkowe, zastosowanie ma osobliwość prognostyczna. Choć GOLEM tego nie porusza, ludzkie oraz nadludzkie (najpewniej trans- i stojące ponad nimi postludzkie) inteligencje mogą współistnieć przez jakiś czas w tej samej cywilizacji. Ta niejednorodność może być efektem na przykład czasochłonności lub innych logistycznych aspektów

samego procesu augmentacji; może stanowić też skutek swoistej umowy społecznej dotyczącej sposobów koegzystencji tych rodzajów bytów. O ile współistnienie tak zróżnicowanych inteligencji będzie miało miejsce i te z wyższego *tier* zmieniałyby rzeczywistość w sposób niezrozumiały — tedy nieprzewidywalny — dla niższych, to nastąpiłby ów wariant osobliwości technologicznej.

Czego GOLEM również nie uwzględnia, to możliwość przecięcia lub pokrycia się prognozowanych ścieżek. Pozaludzka nadludzka inteligencja z pierwszej z nich może wywyższyć ludzkość do stanu postludzkiej nadludzkiej inteligencji (Kurzweil, 2013: 40–41), a w efekcie tego nawet zrównać ją ze swoją lub uczynić jeszcze inteligentniejszą. Z kolei postludzkie nadludzkie inteligencje z drugiej trajektorii mogą wprowadzić pozaludzką nadludzką inteligencję, w tym równą im lub na poziomie nadpostludzkim. Urzeczywistnienie takich scenariuszy ostatecznie zależy od fizycznych (w tym substratowych) ograniczeń samego potęgowania inteligencji, które to ograniczenia zresztą są przedmiotem rozważań GOLEMA (Gomułka & „Preppikoma" Palm, 2021: 359–364); samo potęgowanie w ten sposób pozostaje jednak zgodne z Goodowskim zamysłem eksplozji inteligencji. W każdym przypadku zejścia się bajkowych ścieżek obowiązują te same lub analogiczne kryteria nastania danego wariantu osobliwości technologicznej co w przypadku pojedynczej trajektorii, o ile dana osobliwość nie nastąpiła przed ich zetknięciem.

PODSUMOWANIE

Choć termin „osobliwość technologiczna" po raz pierwszy pojawia się u Vingego, to niemal tą samą nazwą (bez przydawki klasyfikującej) wcześniej posłużył się (wg Ulama) von Neumann w celu opisania nieodwracalnej zmiany ludzkiego życia spowodowanej tempem rozwoju technologii; w międzyczasie Good wprowadził ideę eksplozji inteligencji, tj. tworzenia przez inteligentne istoty bytów inteligentniejszych od nich samych. Echa tych koncepcji oraz astrofizycznych, językowych i matematycznych analogii są dostrzegalne w ujęciu Vingego.

Zaproponowane w tym utworze warianty osobliwości technologicznej nie wyczerpują stanowisk autorów, na bazie których je opracowano. Propozycje te zazębiają się z częścią tych obecnych w typologiach Sandberga i Yudkowskiego, które to typologie można uznać za najbardziej reprezentatywne, lecz są wystarczająco zniuansowane i rozwijają owe klasyfikacje. Co ważne, poprzez użycie słowa „osobliwość" w swojej konstrukcji sygnalizują pokrewieństwo z ogólnym terminem i zmniejszają ryzyko ekwiwokacji. Wszystkie za warunek wstępny uznają samo pojawienie się nadludzkiej inteligencji, ale dopiero jej działania determinują, który wariant się urzeczywistnia.

Związana z Vingem osobliwość prognostyczna zakłada brak wystarczających zdolności ludzi do pojmowania rzeczywistości zmieniającej się na skutek

sprawstwa inteligencji wyższego poziomu; brak rozumienia oznacza, że przewidywanie zmian stanie się co najmniej ograniczone, o ile w ogóle skuteczne.

Oparta na propozycji Kurzweila osobliwość konwergencyjna sprowadza się do ekspansywnego zbiegu procesów potęgowania inteligencji niezależnie od jej substratu, w efekcie dążąc do przemiany wszechświata w komputronium.

Opracowana za Bostromem osobliwość egzystencjalna to wystąpienie zagrożenia egzystencjalnego — unicestwienie pomniejszych inteligencji lub ich zdolności rozwojowych — na skutek procesów związanych z eksplozją inteligencji.

W obu GOLEMOWSKICH trajektoriach — niezależnie od powodu niehomogeniczności inteligencji — ewentualne niepojmowalne działania wyższego *tier* pozostawałyby nieprzewidywalne dla niższego. Jeżeli augmentacja obejmowałaby wszystkie podmioty z danego poziomu w takich okolicznościach lub tak krótkim czasie, że owe aktywności by nie nastąpiły, lub wyższe inteligencje z dowolnego pojmowalnego dla niższych powodu zdecydowałyby się nie działać w ów niepojmowalny dla tych ostatnich sposób, to nie miano by do czynienia z osobliwością prognostyczną. Zagrożenie egzystencjalne może nastąpić w wyniku celowych ograniczeń potęgowania inteligencji lub z powodu nieskuteczności technologii do niej koniecznej; w GOLEMOWSKICH ścieżkach jednak brak antagonizmu sugerującego wprowadzanie takich rozwiązań czy dążenie do unicestwienia pomniejszych bytów.

Te same ustalenia i definicje można uogólnić i zastosować dla porównań dowolnych poziomów czy form inteligencji, nie tylko ludzkich–nadludzkich, ludzkich–postludzkich, ludzkich–pozaludzkich itd. Choć twierdzenia o interakcjach tak zróżnicowanych bytów i sprawstwie wyższego *tier* są w niektórych interpretacjach osobliwości kontestowane jako nieprognozowalne z niższego (Yudkowsky, 2007), to wciąż „podobnie jak potrafimy za pomocą myślenia koncepcyjnego wnioskować odnośnie do charakteru czarnych dziur, choć nigdy nie byliśmy w żadnej z nich, nasza obecna zdolność wnioskowania pozwala na wgląd w następstwa Osobliwości" (Kurzweil, 2013: 474), przy czym tu rozważano następstwa *explicite* zaistnienia nadludzkiej inteligencji.

Raz jeszcze: samo jej pojawienie się nie musi oznaczać osobliwości technologicznej. Hipotetycznym przykładem takiej introdukcji, która nią nie skutkuje, jest historia samego GOLEMA, który to ani nie stanowi zagrożenia egzystencjalnego, ani nie dąży do wzbudzenia inteligencji w całym wszechświecie, ani — potencjalnie — nie zmienia rzeczywistości w niepojmowalny dla człowieka sposób; to ostatnie pozostaje jednak otwarte na interpretację, przez co osobliwość prognostyczna mogła w niej nastąpić (Lem, 2009: 308–310, 313–316, 318–320). Losy i plany tej postaci skupionej na pełnym poznaniu uniwersum stanowią za to interesujący alternatywny scenariusz — taki, w którym następuje *exodus* wyższych inteligencji przynajmniej z cywilizacji niższych, dla których bezpośrednią konsekwencją jest przede wszystkim inspiracja.

BIBLIOGRAFIA

Amato, I. (1991). Speculating in precious computronium. *Science*, *253*(5022), 856–857. DOI: 10.1126/science.253.5022.856.

b.d. (2013). Future trajectories: Singularity (s. 361–364). W: M. More & N. Vita-More (Red.). *The transhumanist reader: Classical and contemporary essays on the science, technology, and philosophy of the human future*. Chichester: Wiley-Blackwell.

Bostrom, N. (2016). *Superinteligencja. Scenariusze, strategie, zagrożenia*. (Przeł. D. Konowrocka-Sawa). Gliwice: Wydawnictwo Helion / Onepress.

Brin, D., Broderick, D., Bostrom, N., „Sasha" Chislenko, A., Hanson, R., More, M., Nielsen, M., & Sandberg, A. (2013). A critical discussion of Vinge's singularity concept (s. 396–417). W: M. More & N. Vita-More (Red.). *The transhumanist reader: Classical and contemporary essays on the science, technology, and philosophy of the human future*. Chichester: Wiley-Blackwell.

Gomułka, J. & „Preppikoma" Palm, J. (2021). GOLEM XIV i hierarchia topozoficzna (s. 355–366). W: F. Kobiela & J. Gomułka (Red.). *Filozoficzny Lem. Wybór tekstów Stanisława Lema i opracowania*. T. 1: *Naturalne czy Sztuczne? Byt, umysł, twórczość*. Warszawa: Wydawnictwo Aletheia.

Gomułka, J. (W przygotowaniu). Golem XIV, czyli o społecznych skutkach kontaktu z Obcym. W: R. Palm, J. Gomułka, K. Grabowska-Derlatka & A. Płoskonka (Red.). *PhilosophyPulp: Volume I*. Kraków: Wydawnictwo Naukowe Uniwersytetu Pedagogicznego.

Good, I.J. (1966). Speculations concerning the first ultraintelligent machine. *Advances in Computers*, *6*, 31–88. DOI: 10.1016/s0065-2458(08)60418-0.

Grygiel, W.P. (2014). *Stephena Hawkinga i Rogera Penrose'a spór o rzeczywistość*. Kraków: Copernicus Center Press.

Heller, M. (1991). *Osobliwy Wszechświat. Wstęp do teorii klasycznej osobliwości kosmologicznej*. Warszawa: Państwowe Wydawnictwo Naukowe.

Institute for the Future. (b.d.). *Artifacts from the future: Tangible, concrete, experiential*. Dostęp: https://www.iftf.org/what-we-do/artifacts-from-the-future/ (06.04.2022).

Jaźniewicz, W. (2021). Bibliografia światowa dzieł Stanisława Lema omawianych w tomie 1 zbioru *Filozoficzny Lem* (s. 537–555). W: F. Kobiela & J. Gomułka (Red.). *Filozoficzny Lem. Wybór tekstów Stanisława Lema i opracowania*. T. 1: *Naturalne czy Sztuczne? Byt, umysł, twórczość*. Warszawa: Wydawnictwo Aletheia.

Keats, J. (2011). *Virtual words: Language on the edge of science and technology*. New York: Oxford University Press.

Kosiński, W. (1981). *Wstęp do teorii osobliwości pola i analizy fal*. Warszawa: Państwowe Wydawnictwo Naukowe.

Krüger, O. (2021). Virtual immortality: God, evolution, and the singularity in post- and transhumanism. Bielefeld: transcript Verlag.

Kurzweil, R. (2013). *Nadchodzi osobliwość. Kiedy człowiek przekroczy granice biologii*. (Przeł. E. Chodkowska & A. Nowosielska). Warszawa: Kurhaus Publishing.

Lem, S. (2009). Golem XIV (s. 207–323). W: S. Lem. *Biblioteka XXI wieku. Golem XIV*. Warszawa: Agora.

Losos, J.B. (2017). *Improbable destinies: Fate, chance, and the future of evolution*. New York: Riverhead Books.

Maj, K.M. (2014). Eutopie i dystopie. Typologia narracji utopijnych z perspektywy filozoficznoliterackiej. *Ruch Literacki*, *60*, *2*(323), 153–174.

Merriam-Webster. (b.d.). *Singularity*. Dostęp: https://www.merriam-webster.com/dictionary/singularity (16.02.2023).

Orliński, W. (2021). *Sieci neuronowe nie wyślą przeciw nam robotów-zabójców. One zaczną na nas wpływać*. Dostęp: https://wyborcza.pl/duzyformat/7,127290,26782071,sieci-neuronowe-nie-wysla-przeciw-nam-robotow-zabojcow-one.html (16.02.2023).

Owczarek, K. (2019). Rozum wyzwolony. *Golem XIV* jako przykład osobliwości technologicznej. *Internetowy Magazyn Filozoficzny HYBRIS*, *46*(3), 69–86. DOI: 10.18778/1689-4286.46.05.

Oxford dictionaries. (b.d.). *Singularity*. Dostęp: https://en.oxforddictionaries.com/definition/singularity (27.06.2017).

Piekutowski, P.F. (2021). Przyszłość przeszłości. Retrotopia w narracjach sztucznej inteligencji Stanisława Lema i Łukasza Zawady. *Przegląd Kulturoznawczy*, *3*(49), 539–553. DOI: 10.4467/20843860PK.21.037.14357.

„Preppikoma" Palm, R. (2022). Teogonia technologiczna. Nominalistyczna koncepcja bóstwa dla transhumanizmu i posthumanizmu (s. 129–143). W: K. Grabowska-Derlatka, J. Gomułka & R. „Preppikoma" Palm (Red.). *PhilosophyPulp: Vol. 2*. Kraków: Wydawnictwo Libron.

Raju, S. (2022). Lessons from the information paradox. *Physics Reports*, *943*, 1–80. DOI: 10.1016/j.physrep.2021.10.001.

Sandberg, A. (2013). An overview of models of technological singularity (s. 376–394). W: M. More & N. Vita-More (Red.). *The transhumanist reader: Classical and contemporary essays on the science, technology, and philosophy of the human future*. Chichester: Wiley-Blackwell.

„Sen" Diaz, A.S. (2008). Eloi. *Hob*, *22*. Dostęp: http://dresdencodak.com/2008/06/07/eloi/ (02.02.2023).

Słownik języka polskiego PWN. (b.d.). *Osobliwość*. Dostęp: http://sjp.pwn.pl/sjp/osobliwosc;2496546.html (16.02.2023).

Ulam, S. (1958). John von Neumann 1903–1957. *Bulletin of the American Mathematical Society*, *64*(3.P2), 1–49. DOI: 10.1090/S0002-9904-1958-10189-5.

Vinge, V. (1983). First word. *Omni*, *5*(4), 10.

Vinge, V. (1986). *Marooned in realtime*. New York: Bluejay Books.

Vinge, V. (1993). *The coming technological singularity: How to survive in the Post-Human era*. Dostęp: http://edoras.sdsu.edu/-vinge/misc/singularity.html (02.02.2023).

Yudkowsky, E.S. (2007). *Three major singularity schools*. Dostęp: https://www.yudkowsky.net/singularity/schools (02.02.2023).