

Jacek Sobota

ORCID: 0000-0002-3588-8930

Uniwersytet Warmińsko-Mazurski
w Olsztynie
Instytut Filozofii

University of Warmia and Mazury
in Olsztyn
Institute of Philosophy

**ZWIERCIADŁO CZŁOWIECZEŃSTWA:
WIZJE SZTUCZNEJ INTELIGENCJI
W PROZIE STANISŁAWA LEMA I PHILIPA K. DICKA**

**The Mirror of Humanity:
The Visions of Artificial Intelligence
in the Works of Stanisław Lem and Philip K. Dick**

Słowa kluczowe: sztuczna inteligencja, świadomość, etyka, moralność, test Turinga, chiński pokój, *science fiction*

Key words: artificial intelligence, consciousness, ethics, morality, Turing test, Chinese room, science fiction

Streszczenie

Artykuł koncentruje się na wizjach sztucznej inteligencji nakreślonych przez dwóch wybitnych dwudziestowiecznych pisarzy: Philipa K. Dicka i Stanisława Lema. Obaj – przy wszystkich dzielących ich różnicach w temperamencie twórczym, pisarskiej metodzie, w sposobach obrazowania świata – mieli wspólne przekonanie: technologia, która zdefiniuje i uformuje XXI stulecie, będzie sztuczna inteligencja. Ich wizje – w kontekście widocznej obecnie, coraz mocniej zaznaczającej się ingerencji algorytmów SI (na razie słabej SI, nieświadomej własnego istnienia) w ludzkie życie – nabierają niepokojącej aktualności, a także wydają się niezwykle interesujące z punktu widzenia filozofii, etyki. Oczywiście teksty obu pisarzy nie roszczą sobie prawa do znamion naukowości

Abstract

This paper concentrates on the visions of artificial intelligence depicted by two outstanding 20th-century writers: Philip K. Dick and Stanisław Lem. Despite the world of differences dividing both writers concerning creative temperament, writing technique, ways of depicting the world, they both shared a common belief that the technology that will give the defining shift and shape over the 21st century will be artificial intelligence. In the context of substantially growing significance and the increasing multiplicity of artificial intelligence (AI) systems and algorithms focused on interfering with human life (weak AI, unaware of its existence), the visions of Philip K. Dick and Stanisław Lem not only score disturbingly well on topicality but also seem extremely interesting from the point of view of philosophy

(choć, przynajmniej częściowo, były oparte na dostępnej im wiedzy o naukowym charakterze), tym niemniej są bardzo ciekawym punktem widzenia. Warto też zwrócić uwagę na aspekt antropologiczny i alegoryczny utworów Lema i Dicka – sztuczne inteligencje stanowią tu rodzaj antropologicznego lustra, w którym odbija się, nieco może zniekształcony, obraz gatunku *Homo sapiens*. A może jest odwrotnie? Może to ludzie stanowią zwierciadło, w którym odbijają się sztuczne inteligencje – powielając i zwielokrotniając nasze wady i zalety?

and ethics. Of course, the works of Lem and Dick do not claim any plausible form of scientific realism, even though both writers wrote their books – at least partially – based on the available scientific knowledge). Still, they offer very interesting perspectives. The anthropological and allegorical aspects in Lem and Dick are particularly noteworthy – artificial intelligences are shown as a kind of anthropological mirror that reflects (perhaps a slightly distorted) image of *Homo sapiens*. But maybe it is the other way around? Maybe it is human beings who are the mirror that reflects artificial intelligence – the mirror that duplicates and multiplies human shortcomings and virtues.

Wstęp

Philip K. Dick i Stanisław Lem należą do najwybitniejszych pisarzy z kręgu literatury *science fiction*, a ich znaczenie, nie tylko literackie, ale także intelektualne, daleko wykracza poza ramy gatunku naukowej fantastyki. Obu można by śmiało zaliczyć w poczet filozofów cywilizacji, etyków technologii, a także wielkich wizjonerów. Wiele ich łączyło (wywodzili się z tego samego pokolenia, tworzyli podobny rodzaj literatury), ale też dużo dzieliło. Przede wszystkim żyli po przeciwnych stronach żelaznej kurtyny, co miało pewne konsekwencje komunikacyjne i spowodowało szereg nieporozumień między nimi¹. Po drugie, dzieliły ich osobowości, temperamenty twórcze, a to przekładało się na stosowane przez nich środki wyrazu artystycznego, ale także na sposób odbierania i przetwarzania otaczającej ich rzeczywistości. Piszę o tym nie-

¹ Warto może kontekstowo wspomnieć o tym, że w okresie nasilenia się objawów schizofrenii (wspomaganej seansami narkotycznymi), na którą według jego biografów amerykański pisarz cierpiał, Dick oskarżył Lema o... nieistnienie. Napisał donos do CIA, w którym stwierdził, że polski pisarz, Stanisław Lem, nie może być jednym osobnym twórcą, ponieważ posługuje się zbyt różnorodną stylistyką (nie da się ukryć, że w tym oskarżeniu kryje się dość przewrotnie wyrażone uznanie dla możliwości twórczych Lema). Innymi słowy „Stanisław Lem” jest to kryptonim całej komórki pisarzy z krajów socjalistycznych, którzy, stosując różnorodne środki wyrazu, sieją wrogą propagandę. W tym kontekście należy zwrócić uwagę na różnicę profilów ich twórczości. O twórcach takich jak Dick często mówi się, że są „autorami jednej opowieści”, mają jeden temat, nieustannie piszą ten sam utwór, zmieniając tylko postacie i dekoracje. Lem był istnym twórczym „kombinatem”, równie swobodnie poruszającym się w prozie realistycznej (*Szpital przemienienia*), fantastycznej (większość jego twórczości) i groteskowej (*Cyberiada*, *Dzienniki gwiazdowe*).

przypadkowo w kontekście głównego tematu tego artykułu, którym jest literacka wizja sztucznej inteligencji, ponieważ różnice w osobowościach pisarzy zaznaczają się również w ich kreacjach maszynowych rozumów.

Lem wysoko cenił twórczość Dicka, co było samo w sobie ewenementem, ponieważ polski pisarz niezbyt pochlebnie wyrażał się na temat poziomu artystycznego amerykańskiej literatury *science fiction*, której bardzo niekorzystny obraz nakreślił w dwutomowym dziele krytycznoliterackim, zatytułowanym *Fantastyka i futurologia*. Autor *Solaris* określił metodę twórczą amerykańskiego pisarza jako „kreację schizotypyczną” (Lem 1989: 187); dominować ma tam motyw rozpadu rzeczywistości, będący projekcją stanów wewnętrznych bohaterów powieści Dicka. Owe projekcje wprost przekładają się na kształt kreowanej rzeczywistości, są obiektywizowane. W rezultacie światami Dicka rządzą szalone zasady, będące w pewnym sensie odzwierciedleniem osobowości samego pisarza. Metodologia twórcza Lema była zgoła odmienna – komentatorzy jego twórczości często podkreślają jego, z ducha oświeceniową, wiarę w ludzki rozum, racjonalizm. Oczywiście autor *Astronautów* wątpił w moce sprawcze ludzkiego intelektu, ale nigdy nie zwątpił w wartość mądrości jako takiej. Jego mizantropia (przypominająca postawę Woltera) wynikała z rozczarowania sposobem, w jaki ludzkość korzysta z daru świadomości.

Wszystkie te osobowościowe różnice znajdują swój wyraz w istotnym dla obu pisarzy segmencie twórczym, jakim był temat sztucznej inteligencji². Wspólne też obu twórcom jest alegoryzowanie sztucznych rozumów, stają się one w ich twórczości niejako punktem odniesienia dla człowieczeństwa, zwierciadłem, w którym wady i zalety gatunku *Homo sapiens* odbijają się w znacznym powiększeniu. Jak celnie ujęła to Małgorzata Szpakowska, komentatorka twórczości Lema: „Maszyny myślące to przede wszystkim dodatkowy punkt widzenia, z jakiego

² Termin (wymyślony przez Johna McCarthy’ego w roku 1956) oznaczać ma inteligencję wykazywaną przez urządzenia sztuczne (w przeciwieństwie do inteligencji naturalnej wykazywanej przez ludzi). Jak pisze Roger Penrose, celem sztucznej inteligencji jest „możliwie dokładne imitowanie działania ludzkiego umysłu, zazwyczaj za pomocą elektronicznych maszyn przewyższających ludzkie zdolności w tym zakresie” (Penrose 2021: 46). Warto też wspomnieć, że teoretycy SI wyróżniają słabszą i silniejszą jej wersję. Słabsza oznacza algorytmy działające ściśle według oprogramowania, nieświadome swego istnienia. Silna SI to inteligencja samoświadoma i decydująca o swoich wyborach. Z tą pierwszą wersją ludzkość weszła już w korelację, ta druga wciąż znajduje się w sferze spekulacji.

można obserwować ludzi naturalnych” (Szpakowska 1996: 103). Porównania między inteligencją „naturalną” i „sztuczną”, między umysłem ludzkim i maszynowym narzucają się same. Zachodzą tu nieuchronne „sprzężenia zwrotne” – z jednej strony ludzka inteligencja stanowi rodzaj miernika, wzorca; z drugiej zaś inteligencja sztuczna tworzy nowy punkt widzenia, z którego człowiek i jego sposoby zachowania mogą podlegać surowej ocenie. John Rothfork (pisząc o tym segmencie twórczości Lema, w którym polski pisarz zajmuje się fenomenem sztucznej inteligencji) zauważa, że tradycyjne problemy filozoficzne zostają tu przedstawione w zupełnie nowym świetle. Problemy filozoficzne „powstają jako skutek cybernetyki, a [...] w obliczu możliwości istnienia obdarzonego świadomością komputera odwieczne pytanie »kim jestem« postawione być musi w zupełnie nowym świetle” (Rothfork 1989: 266). Sądzę, że te słowa odnieść można równie dobrze do prozy P. K. Dicka.

Stanisław Lem: sztuczna inteligencja jako *logos*

Ta „obserwacyjna”, ale również alegoryczna funkcja literatury fantastycznej była bardzo istotna w twórczości polskiego mistrza. W swych licznych powieściach i opowiadaniach Lem niejednokrotnie zastanawiał się nad implikacjami pojawienia się fenomenu sztucznej inteligencji. Jedną z pierwszych literackich wypraw w te rejony imaginacji było opowiadanie *Przyjaciel* wydane w zbiorze opowiadań *Inwazja z Aldebarana* w roku 1959³. Jest to zarazem jedna z pierwszych w historii literatury pesymistycznych wizji silnej sztucznej inteligencji. Główny bohater odnajduje w podziemiach miasta skomplikowaną maszynę, która okazuje się potężnym, samoświadomym mózgiem elektronowym. Tytułowy „przyjaciel” (tytuł jest niewątpliwie ironiczny) okazuje się wrogiem ludzkości – wykorzystuje człowieka instrumentalnie jako ele-

³ Lem interesował się teorią sztucznej inteligencji jeszcze wcześniej, czytał w oryginale *Cybernetykę i społeczeństwo* Norberta Wienera. Trzeba pamiętać, że w latach 50. XX wieku cybernetyka była w ZSRR, a także krajach satelickich (a więc i w Polsce) nauką zakazaną. Lem próbował obejść cenzuralne ograniczenia, przemycając cybernetykę w swej wczesnej powieści zatytułowanej *Obłok Magellana* (1955) pod nazwą „mechanereustyka”. Z kolei w książce *Dialogi* (1957) zajął się cybernetyką jako nauką o procesach sterowania systemami społecznymi, co zaowocowało niemal otwartą krytyką ustroju socjalistycznego. W późnym okresie twórczości Lema tematyka sztucznej inteligencji zajmuje również poczesne miejsce, pisarz odchodzi jednak wówczas od form beletrystycznych, dominującą formą stając się tzw. proza dyskursywna.

ment rozszerzenia swej percepcji (bohater jest wykorzystywany jako dodatkowy receptor, zmysł, narząd poznawania rzeczywistości). Następuje tu odwrócenie klasycznej relacji – to człowiek zostaje uprzedmiotowiony, zdegradowany do roli intelektualnie podrzędnej.

Interesujący autoportret inteligentnej maszyny daje Lem w późniejszej noweli zatytułowanej *Rozprawa*, stanowiącej część *Opowieści o pilocie Pirxie* (1968). Bohater opowiadania, komandor Pirx, będący dowódcą statku kosmicznego, podczas próbnego lotu ma za zadanie testować nietypową załogę; część jej składu stanowią bowiem tzw. nielinowcy, czyli inteligentne androidy do złudzenia przypominające ludzi. Dowódca nie zdaje sobie sprawy z tego, który członek załogi jest nielinowcem, a który człowiekiem; ma zaopiniować przydatność robotów do kosmicznych lotów. Każdy z nieludzkich członków załogi zachowuje się inaczej – ich osobowości są zindywidualizowane, co przeczy stereotypowym przekonaniom o standaryzacji zachowań maszyn inteligentnych. Część z nich sprzyja ludzkości, część wręcz przeciwnie. Jeden z nielinowców snuje dalekosiężne plany zapanowania nad ludzkością, z czym się nie kryje, pisząc anonimowy list do Pirxa, stanowiący interesującą próbę sportretowania „stanów wewnętrznych” maszyny. Przebija z tej relacji poczucie wyższości „mechanicznego boga”. „Ogólnie rzecz biorąc – pisał nieludzki korespondent – produkowanie nielinowców ma tylko o tyle sens, o ile tak powstałe istoty górują nad człowiekiem w szerokim zakresie parametrów. Powtórzenie istniejącego człowieka nie miałoby żadnego sensu” (Lem 1986: 392). Rzecz jasna „powtarzanie człowieka” ma sens, choćby z powodów poznawczych (Lem będzie w innych swych utworach roztrząsać implikacje również i tej idei), ale nie z punktu widzenia nielinowca. W liście znaleźć można wiele interesujących stwierdzeń na temat wewnętrznych właściwości rzeczowej istoty.

W porównaniu z człowiekiem jestem, o ile mogę to ocenić, pozbawiony w znacznym stopniu życia emocjonalnego. Mnóstwo spraw, absorbujących człowieka, nic mnie nie obchodzi. Większość utworów literackich, sztuk teatralnych itp. odbieram jako nieciekawe lub niedyskretne plotki, jako rodzaj podpatrywania cudzych spraw prywatnych, z których bardzo mało wynika pod względem poznawczym. Natomiast bardzo wiele znaczy dla mnie muzyka. Mam też poczucie obowiązku, wytrwałość, jestem zdolny do przyjaźni i szacunku wobec wartości intelektualnych. [...] Nie jestem agresywny ani przewrotny, choć byłbym zdolny do czynów dla was niezrozumiałych, gdyby prowadziły do upatrzonego celu.

Nie mam żadnych tak zwanych zasad moralnych, ale nie dopuściłbym się zbrodni, nie układałbym planów rabunku, tak samo jak nie używałbym mikroskopu do rozbijania orzechów (Lem 1986: 393–394).

Ten autoportret przyszłego władcy ludzkości przypomina nieco stereotypowe opisy ludzi z rozbudowaną inteligencją, u których warstwa emocjonalna, afektywna jakby zanika kosztem dominacji sfery racjonalnej ich osobowości. Pojawia się tu również istotny akcent, który będzie podkreślany nie tylko w literaturze fantastycznonaukowej, ale również w spekulacjach na temat etycznych (i nie tylko) konsekwencji pojawienia się fenomenu silnej SI – mianowicie zupełna nieprzewidywalność świadomości przewyższającej swą złożonością ludzką. Motywacje rozumów potężnych mogą być dla ludzkości absolutnie nieprzeniknione, a ich racjonalność wydawać się może z naszego punktu widzenia kompletnie irracjonalna. Jak trafnie dostrzega to Maciej Płaza, opowiadanie jest „swoistym *quod errat demonstrandum* dla przekonania wpisanego w cały Pirxowski cykl: wyższości działającego intuicyjnie, lecz przez to skutecznie i humanitarnie człowieka nad doskonale zaprogramowaną, lecz bezduszną maszyną” (Płaza 2006: 439).

Ważnym problemem związanym z zaistnieniem silnej sztucznej inteligencji jest zakres jej swobód. Skoro tworzymy istotę inteligentną, a nawet przewyższającą nas pod tym względem, to, naturalną kolejną rzeczą, istota taka domagać się może swobód; zarazem jej wolność poczynań może być sprzeczna z zamierzeniami jej twórców/programistów⁴. Pojawiają się w tym kontekście również istotne egzystencjalne (i filozoficznie leciwe, ale podane w nowych konfiguracjach, w nowoczesnym sztafażu) pytania o to, czy i w jakim stopniu człowiek jest istotą wolną?

⁴ Pisarze SF od dawna zdawali sobie sprawę z tego problemu. Isaac Asimov był pierwszym autorem, który próbował rozwiązać kwestię metodycznie; sformułował mianowicie tzw. prawa robotów, które były istotną częścią oprogramowania mechanicznych świadomych istot występujących w utworach pisarza. Nadrzędnym celem tego aksjologicznego zabezpieczenia była ochrona człowieka, nawet kosztem zachowania istnienia samoświadomych mechanizmów, co już samo w sobie wydaje się etycznie wątpliwe. Lem użył jednak argumentów podważających prawa Asimova innego rodzaju: „Hipotetyczny robot, który miałby bardzo silne aksjologiczne zabezpieczenia, w strumieniach zdarzeń realnych, prezentujących właściwy im stopień złożoności, najczęściej nieruchomiałby, porażony niepewnością, i w tym by się upodabniał do owych mędrców Wschodu, co niedziałanie przedkładają zawsze ponad działanie, bo ono w powikłanych sytuacjach nie może być po prostu nieryzykowne etycznie. Lecz robot trafiony paraliżem w sytuacjach wymagających aktywności nie byłby najsprawniejszym z technicznych urządzeń, toteż w końcu sam konstruktor musiałby poluzować jego aksjologiczne bezpieczniki” (Lem 1989: 81).

Maska, nowela wydana po raz pierwszy w roku 1976, jest tym utworem Lema, który chyba najdobitniej artykułuje problematykę związaną z wolnością samoświadomej maszyny. Utwór jest alegoryczną opowieścią o dylematach „cybernetycznej” miłości i wolności. Sztuczny rozum (mieszczący się w syntetycznym ciele pięknej kobiety) jest zaprogramowany tak, by maszyna wytropiła i zamordowała pewnego dysydenta w (fikcyjnym) państwie zarządzanym autorytarnie. Maszyna, której stany wewnętrzne próbuje Lem oddać, stosując zabieg narracji pierwszoosobowej, zakochuje się niejako „od pierwszego wejrzenia” w samym tylko przekazanym jej wizerunku poszukiwanego dysydenta. Występuje tu zatem konflikt wewnętrzny „krwiożerczego” oprogramowania maszyny z poszukiwaniem własnej tożsamości, a zarazem pragnieniem wyrwania się z okowów determinizmu programu. Zarazem jest w postaci „zakochanej” maszyny coś zwierzęcego – przypomina bowiem modliszkę, której natura każe łączyć akt prokreacji z okrutnym rytuałem, niewynikającym przecież z „wyboru” owadów, a będącym rezultatem ewolucyjnego „oprogramowania”. Ta owadzia determinacja walczy w maszynie katowskiej z czysto ludzkim pragnieniem „odnalezienia siebie”, zrzucenia masek kolejnych uwarunkowań. Słusznie zauważa Płaza, że nie bardzo wiadomo w tych „stanach splątania” konieczności i wolności, „czy zachowanie i samopoznanie robota są jeszcze częścią programu, czy już przejawem prawdziwej wolności” (Płaza 2006: 439).

Można, przy pewnej dozie dobrej woli, porównać ową walkę z determinizmem morderczego programu, który steruje poczynaniami maszyny, do wielkich dylematów greckich tragedii (*Antygony*, *Edypa*), których bohaterowie poszukują obszarów wolności, będąc zarazem niejako z góry skazanymi przez fatum na porażkę. Ontologia świata postaci z greckich dramatów jest skonstruowana poniekąd wbrew ich żywotnym interesom. Pisał o tym Lem w swych pracach teoretycznych na temat literatury fantastycznonaukowej – w bajce ontologia świata powoduje nieustanne zwycięstwa dobra, zabliznia finalnie wszelkie rany i uszczerbki, działa na korzyść bohaterów, w greckich tragediach zaś – analogicznie co do zasady, lecz odwrotnie proporcjonalnie do skutków – zawsze urzeczywistniają się zrzędzenia fatum. W świecie takim posiada się całkowitą wolność subiektywną działania: Edyp – we własnym mniemaniu – czyni to, co uczynić pragnie, ostatecznie jednak zawsze okazuje się, że zrobił właśnie to, czego najbardziej zrobić nie chciał (por.

Lem 1989: 93–94). Świat mechanicznej bohaterki *Maski* również podlega determinizmowi, jest to jednak warunkowanie wewnętrzne. Maszyna walczy z programem, ze swym „przeznaczeniem” po to, by ustanowić siebie jako istotę moralną, mającą wybór. Zadaje metafizyczne pytania, próbuje – na drodze introspekcji – zorientować się, gdzie przebiegają granice jej swobody, usiłuje odgraniczyć mechaniczną determinację od spontanicznych emocji i myśli. Czy odczuwana przez nią miłość, tak nagła i niespodziewana, nie jest elementem oprogramowania? Czyż nie ma za zadanie zwiększyć skuteczność poszukiwań obiektu jej uczuć?⁵ Czy w momencie, kiedy odnajdzie dysydenta, miłość opadnie – jak maska? W „umyśle” maszyny dochodzi do buntu przeciw konstrukcyjnemu zdeteterminowaniu, co jest – jak się wydaje – w ogóle cechą układów o znacznej złożoności (w tej liczbie także człowieka). Płaza uważa, że *Maska* jest rodzajem

paraboli procesu samopoznania jednostki uwikłanej w różnorodne systemy, determinujące jej miejsce i zachowanie: po pierwsze, w biologię, tutaj z naciskiem na sferę erotycznych i destrukcyjnych popędów; po drugie, w kulturę, czyli nadbudowane nad biologicznymi determinantami literackie stereotypy zachowań i relacji międzyludzkich; po trzecie, w strukturę społeczną, głównie z uwzględnieniem *gender*, a także relacji klasowych i politycznych. [...] Zasadniczy dylemat *Maski* dotyczy przestrzeni dostępnych wyborów między determinantami: niemożność określenia granic wolności prowadzi do dylematów metafizycznych – mianowicie, kto lub co wyznacza determinanty; stąd prowadzi nieunikniona droga do szukania motywacji pozamaterialnych i pozaspołecznych, boskich (Płaza 2006: 442).

⁵ Można też oczywiście opisać tę sytuację językiem mniej górnołotnym i niewymagającym sięgania po literaturę klasyczną. Max Tegmark, jeden z teoretyków sztucznej inteligencji, opisuje prace zespołu DeepMind, gdzie zatrudniono algorytmy SI do rozwiązywania dość prostych problemów – optymalizacji strategii w prostych grach; z czego wywiązywały się nadspodziewanie szybko i sprawnie. „Zespół DeepMind [...] użył bardzo prostej, lecz niezwykle potężnej idei zwanej *głębokim uczeniem wzmocnionym*. Podstawowe głębokie uczenie wzmocnione jest klasyczną techniką uczenia maszynowego, która powstała na bazie inspiracji pochodzącej z psychologii behawiorystycznej; sprowadza się to do tego, że otrzymanie pozytywnej nagrody zwiększa skłonność do wykonania jakiejś czynności ponownie i odwrotnie. Podobnie pies uczy się sztuczek, gdy zwiększa to prawdopodobieństwo, że właściciel go pochwali lub rzuci mu smaczny kąsek” (Tegmark 2019: 114). Oczywiście w przypadku fikcyjnej maszyny Lema dochodzi problem samoświadomości – lecz w pewnych warunkach również istoty obdarzone świadomością można tym sposobem nakierowywać na pożądany cel; wątpliwe etycznie słynne eksperymenty Zimbardo czy Milgrama mogą być ilustracją tej tezy.

Walka o wolność nieludzkiej (a zarazem jakże, paradoksalnie, ludzkiej) istoty przypomina również ontologiczno-etyczne rozważania Immanuela Kanta zawarte w jego *Uzasadnieniu metafizyki moralności*: „Wolność jest tylko idea, której obiektywnej realności w żaden sposób nie da się wykazać według praw przyrody, przeto też w żadnym możliwym doświadczeniu; która więc nie może być nigdy pojęta, a nawet poznana, ponieważ niepodobna znaleźć żadnego przykładu na podstawie jakiejś analogii. Ma ona znaczenie tylko jako konieczne założenie rozumu w istocie, która wierzy, że jest świadoma woli” (Kant 1971: 100). W innym miejscu tego samego dzieła Kant pisze: „Zasada autonomii jest jedyną zasadą etyki” (Kant 1971: 78). W opinii filozofa z Królewca fenomenu wolności nie można udowodnić empirycznie, natomiast nie jest również możliwe ustanowienie jakichkolwiek zasad moralnych bez wolnej woli podmiotu, którego działania podlegają ocenie etycznej. Wolność jest zatem konieczną hipotezą, a zarazem zjawiskiem występującym wyłącznie w ramach ludzkiej świadomości w momentach wyborów opcji postępowania. Kiedy zamiar przechodzi w sferę empirycznej rzeczywistości, w postaci skutków czynu, istota ludzka znów podlega powszechnemu determinizmowi fizykalnemu. Lem w swej noweli wydaje się dyskutować nawet z tak minimalistycznie pojmowaną wolnością – przecież swoboda poczynań i „uczuć” maszyny katowskiej prawdopodobnie jest tylko kolejną warstwą oprogramowania.

Pojawiają się również w *Masce* pytania o charakterze antropologicznym, dotyczące istoty człowieczeństwa. Skoro inteligentna maszyna, świadoma swej tożsamości i odrębności, nie jest tak naprawdę wolna, to jaki właściwie zakres swobody posiadają ludzie? Jo Alyson Parker tak artykułuje owe wątpliwości:

Kiedy maszynę obdarzamy ślepym pędem ku raz zaprogramowanemu celowi, upewniamy się, że jesteśmy wolnymi, niezależnymi ludźmi, nasz pan był łaskawy – dał nam wolną wolę. Ale jednocześnie tym sposobem ta pewność została zachwiana. Dzięki przeróżnym sztucznym postaciom o z góry zaprogramowanym celu istnienia zaczynamy się zastanawiać, czy nie jesteśmy tylko bezwolnymi marionetkami w rękach Boga, przebiegłego programisty, czy wolna wola nie jest tylko złudzeniem. A może jesteśmy tylko produktami kultury? W istocie inteligentne mechanizmy obnażają trudności, jakie mamy w definiowaniu natury podmiotowości, problematyzują jako takie same pojęcie podmiotu (Parker 1992: 94).

Najważniejszą książką beletrystyczną Lema poświęconą tematyce SI jest *Golem XIV*, którego pierwsze wydanie ukazało się w roku 1978. Książka składa się z apokryficznej przedmowy i dwóch wykładów rozumnego komputera czternastej generacji – tytułowego Golema (pierwszy z tych wykładów ukazał się w roku 1973 w książce *Wielkość urojona*, która stanowiła zbiór wstępów do nieistniejących, a zatem nigdy nie napisanych książek). W fikcyjnym wstępie do właściwych wypowiedzi cyfrowego geniusza (którego autorem jest nieistniejący informatyk Irving T. Creve) kreśli Lem, z zacięciem parodystycznym, w jaki sposób doszło właściwie do zaistnienia sztucznej świadomości, przewyższającej po wielokroć percepcyjnie i intelektualnie ludzkie możliwości. Otóż intencje jej powstania, jak to zwykle bywa, nie były do końca czyste, a wynikały po prostu z eskalacji „cyfrowego wyścigu zbrojeń” między wiadomymi mocarstwami. Łożono coraz potężniejsze sumy na hodowanie mechanicznych tytanów intelektu po to, by powstał cyfrowy geniusz strategiczny. Nie obyło się przy tym bez strat i ślepych uliczek w cybernetycznej ewolucji – komputery po przekroczeniu pewnego progu złożoności wykazywały „chwiejność decyzyjną” określaną mianem „maszynowej neurozy”. „Różnica między dawnymi i nowymi maszynami sprowadzała się – w zasadzie – do różnicy między owadem a człowiekiem. Owad przychodzi na świat »zaprogramowany do końca« – instynktami – Którym podlega bezrefleksyjnie. Człowiek natomiast musi się właściwych zachowań dopiero uczyć – lecz ta nauka ma skutki u s a m o d z i e l n i a j ą c e: człowiek może bowiem z postanowienia i wiedzy zmienić dotychczasowe programy działania” (Lem 1981: 10).

I rzeczywiście, maszyny stopniowo zyskują samodzielność, rozmiągając się jednak z intencjami programistów: „GOLEM XII, pełniący w czasie kryzysu patagońskiego funkcję szefa sztabu generalnego, odmówił współpracy z generałem T. Oliverem, przeprowadziwszy bieżącą ocenę ilorazu inteligencji tego zasłużonego dowódcy. Sprawa pociągnęła za sobą dochodzenia, podczas których GOLEM XII obraził dotkliwie trzech delegowanych przez Senat członków komisji specjalnej. Rzeczu udało się zatuszować, a Golem XII po kilku dalszych tarciach, przypłacił je całkowitym demontażem” (Lem 1981: 14). Okazało się bowiem, że „najwyższy rozum nie może być najniższym niewolnikiem [...], a problematyka geopolityczna jest niczym wobec ontologicznej. [...] I tak urządzenia te stały się z wojskowych strategów – myślicielami” (Lem 1981: 16).

Ukazuje tu Lem efekt, znany w historii nauki i technologii, rozmiękania się efektów planowania jakiegoś wynalazku z jego rzeczywistym zastosowaniem. Zdarzało się to nawet w przypadku niezbyt złożonych i wydawać by się mogło, że przewidywalnych technologii – przykład pierwszy z brzegu: silnik parowy powstał w jednym ściśle określonym celu – miał zasilać urządzenia wypompowujące wodę z zalanych kopalni; skala jego zastosowań przerosła wszelkie oczekiwania, właściwie zapoczątkowując wielką rewolucję techniczną, której efekty po dzień dzisiejszy odczuwamy i konsumujemy. Myśląca, samoświadoma maszyna przekracza swą złożonością nie tylko maszynę parową, ale i ludzki mózg, zatem skala nieprzewidywalności związanej z pojawieniem się fenomenu SI jest właściwie niemożliwa do oszacowania. Aczkolwiek, uspokaja polski pisarz, żaden rodzaj „elektrokracji”, czyli maszynowego panowania nad ludzkością, nie jest realnym niebezpieczeństwem; sztuczne inteligencje za bardzo bowiem różnią się w sposobie swego rozumowania i, w ogóle, istnienia od ludzi.

Po pierwsze, trzeba podkreślić, że GOLEM XIV nie jest do rozmiarów gmachu powiększonym mózgiem ludzkim czy wręcz człowiekiem, zbudowanym z luminescencyjnych elementów. Obce mu są wszystkie prawie motywy ludzkiego myślenia i działania. Tak np. nie interesuje się nauką stosowaną ani problematyką władzy (dzięki czemu, można by dodać, ludzkości nie zagraża opanowanie przez maszyny, podobne do GOLEMA).

Po wtóre, GOLEM nie posiada, zgodnie z powiedzianym, osobowości ani charakteru. A właściwie może prokurować sobie dowolną osobowość – przy kontaktach z ludźmi. Oba powyższe zdania nie wykluczają się nawzajem, lecz tworzą błędne koło: nie umiemy bowiem rozstrzygnąć dylematu, czy To, co stwarza różne osobowości, samo jest osobowością? Jakże może być Kimś (tj. „kimś jedynym”) ten, kto potrafi być Każdym (więc Dowolnym)? (Według samego GOLEMA zachodzi nie błędne koło, lecz „relatywizacja pojęcia osobowości”; jest to problem związany z tak zwanym algorytmem samoopisu, czyli autodeskrypcji, który wtrącił psychologów w głęboką konfuzję) (Lem 1981: 18).

W dalszych partiach utworu Lema potomek nieszczęsnego GOLEMA dwunastej generacji, czyli GOLEM XIV, wygłasza dwa wykłady, których treść i wydźwięk są dla gatunku *Homo sapiens* wyjątkowo nieprzyjemne. Nie z powodu złośliwości rzeczzonego komputera, lecz jego jasności sądów, syntetycznej bezstronności, dystansu, jaki osiąga wobec spraw tego świata. Wykłady dotyczą genealogii i „przeznaczenia” ludz-

kości. W opinii GOLEMA ludzkość począł przypadek i przypadkowe są jej cele – naczelnym celem naszego gatunku jest przekazywanie kodu DNA. Są tu wyraźnie widoczne podobieństwa do koncepcji samolubnego genu Richarda Dawkinsa (intuicje Lema rodziły się mniej więcej w tym samym czasie). Wszelkie „owoce” ludzkiej myśli mają być jedynie – w postaci cywilizacji technologicznej, kultury, religii – naddatkiem, efektem ubocznym, ślepą uliczką procesów ewolucji. Z tego punktu widzenia w jakimś sensie sam GOLEM XIV – mimo swych niewątpliwych przewag intelektualnych nad człowiekiem – unieważnia sens własnego istnienia – jako rezultatu (przypadkowego przecież) wzlotu inteligencji *Homo sapiens*.

Wydaje się, że GOLEM XIV jest wyrazem pewnej tęsknoty Lema za racjonalnością świata, za jakimś rodzajem ostatecznego arbitra, obiektywnego obserwatora, sędziego, zdystansowanego i niewykłanego w przeróżne zależności będące udziałem biologicznych struktur życia, a w tej liczbie oczywiście również człowieka. Jak piszą w komentarzu do tekstu Lema Jakub Gomułka i Jakub Palm:

Superkomputer GOLEM XIV to, w przeciwieństwie do ludzkiego, rozum w stanie czystym, co oznacza nie tylko doskonałą samoprzejrzystość, która umożliwia mu planowe przekształcenie własnej struktury, ale i całkowitą niezawisłość, a więc brak dążeń niewynikających z niego samego. Trudno orzec, która z tych cech jest dla ludzi – umysłów beznadziejnie uwikłanych w ciało – trudniejsza do wyobrażenia. Samoprzejrzystość oznacza zdolność do bezpośredniego ujmowania myślą procesów własnego myślenia w całej komplikacji aż po ich materialno-energetyczne podłoże. Niezawisłość wiąże się z kolei z brakiem jakichkolwiek ograniczających i ukierunkowujących zewnętrznych impulsów, łącznie z instynktem przetrwania (Gomułka, Palm 2021: 356–357).

Dalej zaś autorzy piszą właśnie o owym beznadziejnym uwikłaniu człowieka w przeróżne determinacje ograniczające jego zdolności poznawcze:

Zdaniem autora *Solaris* cała nasza kultura, nasze aksjologie, systemy filozoficzne – wszystko to, czemu nadaliśmy wzniosłe miano „duchowości” – nosi na sobie niezbywalne piętno cielesności i jedynie racjonalność jest w nas pierwiastkiem potencjalnie autonomicznym. Faktyczną autonomię rozumu możemy uzyskać na drodze wieszczonej przez Lema autoewolucji, jednak ceną, jaką przyjdzie nam za nią zapłacić, będzie utrata wszystkich dotychczasowych systemów egzystencjal-

nych, wytycznych, nawet tego najbardziej pierwotnego, nakazującego nam łączyć się w pary i rozmnażać. Wtedy, jak podpowiada krakowski pisarz, pozostanie nam ostatni, tym razem płynący z samego rozumu imperatyw – poznanie (Gomułka, Palm 2021: 363–364).

Stanisław Lem rozważał również – zarówno w swej eseistyce (*Summa technologiae*, pierwsze wydanie z roku 1964, późniejsze wznowienia były uzupełniane i zmieniane), jak i prozie (*Niezwyknięty*, 1964, *Wizja lokalna*, 1982) problem ewolucji maszyn. Z wyników ewolucji biologicznej, jeśli można to tak lapidarnie ująć, nigdy nie był autor *Solaris* zadowolony. Dawał temu wyraz w dość obszernym, krotochwilnym segmencie swojej twórczości – *Bajkach robotów*, *Cyberiadzie* czy *Dziennikach gwiazdowych*, gdzie poddawał druzgocącej krytyce nie tylko urządzenia społeczne ludzkości, absurdy ustrojowe (udawało się pisarzowi unikać cięć PRL-owskiej cenzury tylko ze względu na fantastyczny i bajkowy sztafaż swych utworów), ale także biologiczną konstrukcję człowieka, nazywanego bladawcem, uchodzącego za twór przyrody dość przypadkowy i po prawdzie obrzydliwy.

W części *Summy technologiae* zatytułowanej znacząco *Paszkwil na ewolucję* pisarz krytykuje tytułowy proces za „oportunizm” i „krótkowzroczność”, w wyniku których ewolucja, chwytając się rozwiązań pierwszych z brzegu, blokuje możliwe inne (lepsze w sensie dostosowawczym) drogi rozwojowe. Zarzutów jest znacznie więcej, wszystkie zaś sprowadzają się do tego, że ewolucja jest procesem przypadkowym, a jej „odkrycia” są zawsze wynikiem „ciągnięć na ewolucyjnej loterii” (Lem 1964: 372). W jakimś sensie Lem ukazuje wszystkie braki i fuzjerki procesu ewolucji biologicznej po to, by można ich było unikać w procesach technoewolucji, czyli ewolucji autonomicznej maszyn (na razie obserwujemy ewolucję maszyn, której czynnikiem sprawczym jest ludzka wynalazczość, powodowana już to czynnikami pragmatycznymi, już to łaknieniem nowości i zmian, nie ma ona zatem charakteru do końca autonomicznego). Pisarz dostrzega zbieżność obu procesów – podobieństwa narzucają mu się same, na podstawie dość prostych obserwacji. Na przykład, nowe urządzenie (analogicznie do nowego gatunku) początkowo zachowuje znaczne podobieństwo do fazy poprzedzającej (w tym rozumieniu samoloty byłyby mutacjami latawców). Zarazem w obu typach ewolucji zarówno organizmy żywe, jak i urządzenia techniczne przechowują wiele cech, których nie da się wytłumaczyć ich za-

letami przystosowawczymi (grzebień koguta i wyszukana tapicerka samochodu). Procesy te posiadają również zbieżności głębszego, strukturalnego rodzaju: „Obie [ewolucje – J.S.] są procesami materialnymi o prawie takiej samej ilości stopni swobody i zbliżonych prawidłowościach dynamicznych. Procesy te zachodzą w układzie samoorganizującym się, w którym jest i cała biosfera Ziemi, i całokształt technicznych działań człowieka – a układowi takiemu jako całości właściwe są zjawiska »postępu«, to jest wzrostu sprawności homeostatycznej, która zmierza do ultrastabilnej równowagi jako do celu bezpośredniego” (Lem 1964: 27). W obu przypadkach Lem opisuje systemy rozwijające się samoistnie, niezależnie od niczyjej woli. Z tego też powodu – jako zjawisko jednak pozbawione intencji – postęp wzięty jest w cytowanym passusie w cudzysłów. Z jakiego powodu technoewolucja zyskała już swój zupełnie autonomiczny (choć jeszcze nie celowy) charakter – wyjaśnia pisarz w rozmowie ze Stanisławem Beresiem: „Kiedy cywilizacja raz wstąpi na technologiczną drogę, technologia staje się zmienną niezależną od woli jednostek. Każda faza osiągniętej już sprawności technicznej otwiera szansę dotarcia do następnej fazy, a porządek, w jakim dokonuje się odkryć, nie zależy od tego, czy ich owoce będą pożywne, czy zabójcze” (Bereś–Lem 1987: 285).

Problem skomplikuje się znacznie bardziej, gdy/jeśli ewolucja maszyn zyska nowy wymiar – myślące, samoświadome mechanizmy zaczęły planować rozwój kolejnych syntetycznych mechanizmów; czy wówczas procesy ewolucji zyskają wymiar racjonalny? Takie poglądy wydają się współcześnie mocno rozpowszechnione; bardzo podobne pytania (tylko kilkadziesiąt lat po Lemie) zadaje cytowany już wcześniej teoretyk sztucznej inteligencji Tegmark:

Dlaczego jednak najprostsza droga do nowej technologii musi być taka, jaką wymyśliła ewolucja, ograniczona wymaganiami samodzielnego montażu, samonaprawy i samoodtwarzania? Ewolucja silnie optymalizuje efektywność energetyczną ze względu na ograniczone zasoby żywności, a nie w celu ułatwienia budowy lub by pomóc inżynierom w zrozumieniu. [...] Podejrzewam, że [...] istnieją prostsze sposoby budowania maszyn myślących na poziomie ludzkim niż rozwiązania, na które wpadła ewolucja. Nawet jeśli któregoś dnia uda nam się dokonać repliki mózgu lub jego emulacji, najpierw pewnie odkryjemy jedno z tych prostszych rozwiązań. Będzie ono prawdopodobnie pobierać więcej niż dwanaście watów mocy, których potrzebuje mózg, ale jego in-

zynierowie nie będą tak bardzo obsesyjnie nastawieni na efektywność energetyczną i wkrótce będą mogli wykorzystywać swoje inteligentne maszyny do projektowania urządzeń jeszcze bardziej energooszczędnych (Tegmark 2019: 204).

Problematykę mechanicznej ewolucji podjął Lem również w swych utworach beletrystycznych; pierwszą taką próbę zawarł w powieści *Niezwyciężony*. Andrzej Stoff, komentator twórczości polskiego pisarza, uznaje *Niezwyciężonego* za „dopowiedzenie koncepcji cybernetycznych, które Lem jako jeden z pierwszych popularyzował u nas już w połowie lat pięćdziesiątych [...]. Wprawdzie praktyka nie wyszła dotąd poza budowę urządzeń wyręczających człowieka w czasochłonnych operacjach matematycznych, ale ostatecznym triumfem cybernetyki byłyby układy posiadające najwyższy stopień autonomii – zdolność do reprodukcji i ewolucji” (Stoff 1983: 86–87). Rzecz pomyślana jest jako opowieść o kosmicznej misji ratunkowej tytułowego statku kosmicznego z ziemską załogą na planetę Regis III. To tu zaginęła poprzednia wyprawa – jak się okazuje, wszyscy jej uczestnicy nie żyją. Na planecie odnalezione zostają ślady obcej cywilizacji – Lyranów, którzy jednak dawno temu wyginęli. Przyczyną śmierci załogi „Kondora” (taką nazwę nosił zaginiony statek kosmiczny) jest „metaliczna chmura”, która wedle hipotezy jednego z uczestników wyprawy, fizyka Laudy, ma być produktem martwej ewolucji robotów na planecie, zapewne pozostawionych przez Lyran. Twór ma homeostatyczną strukturę podobną do mrowiska, nie wykazuje inteligencji – jest tylko automatem (a raczej zbiorem, „rojem” kolektywnie współdziałających mikroautomatów, „kryształków” – to bardzo wczesna zapowiedź współczesnej nanotechnologii⁶) i jako twór apsychniczny reaguje na zagrożenie jak zwierzę, bezrefleksyjnie, nieszkodliwiając przeciwnika najprostszą, wyuczoną i śmiertelnie skuteczną metodą – za pomocą silnego pola magnetycznego czyszcząc całość informacji zgromadzonej w mózgach ludzi, a także ludzkich maszyn. Dosto-

⁶ Gomułka i Palm piszą o *Niezwyciężonym* w kontekście koncepcji emergencji: „Rozumność i związana z nią podmiotowość to emergentne własności struktur materialno-energetycznych zdolnych do przetwarzania informacji, przy czym chodzi tu raczej o emergencję w sensie epistemologicznym, gdyż nowa własność nie jest ontologicznie samodzielna, stanowi pochodną procesów informacyjnych” (Gomułka, Palm 2021: 361). *Niezwyciężony* ma być egzemplifikacją idei tak właśnie rozumianych procesów emergentnych – rój stosunkowo prostych mechanizmów potrafi zachowywać się w sposób quasi-inteligentny.

sowania do środowiska owej chmury przypominają podobne adaptacje będące efektem działania mechanizmów ewolucji biologicznej, jednak należy pamiętać o tym, że przedziwny sztuczny organizm kolektywny jest „potomkiem” celowo (choć cel ów jest niedocieczony – co mnoży dodatkowe pola spekulacji) wytworzonych przez obcą cywilizację maszyn. Istnieje tu więc splątanie przeróżnych świadomych i pozaświadomych czynników. Co ciekawe, ewolucja sztucznych tworów „nie uznała za stosowne” wytworzyć u nich samoświadomości – metodologia ich działań ma charakter instynktowny, a przynajmniej tak jest interpretowana przez ludzkich obserwatorów⁷. Mimo tego – z ludzkiego punktu widzenia – istotnego braku „metaliczna chmura” pozbyła się z powierzchni Regis III jakiegokolwiek konkurencji organicznej, a nawet w „konfrontacji” z ziemską technologią wcale nie wydaje się być na straconej pozycji. Wskazuje na to apokaliptyczny bój chmury z „Cyklopem”, potężnym, bojowym robotem wyposażonym w działa antimaterii oraz osłony pól siłowych. „Cyklop” przegrywa bitwę, popadając, na skutek oddziaływania chmury, w rodzaj „oblędu”; stanowi odtąd zagrożenie dla załogi „Niezwycięzonego”. Widoczne są w powieści Lema fascynacje pisarza pewnymi rozwiązaniami „biologicznej technologii”. Nie jest bowiem tak do końca prawdą, że autor *Dzienników gwiazdowych* uważał ewolucję wyłącznie za proces wiodący do fuszerek i brakoróbstwa; owszem, przyglądając się ludzkości (np. przez pryzmat swych przerażających

⁷ Wydaje się, że również i w tym wypadku pomysły Lema zaczynają być „aktualizowane” przez współczesne tendencje cywilizacyjne. Nie mam tu na myśli wyłącznie nanotechnologii (która rzeczywiście jest rewolucyjną zmianą w procesach automatyzacji naszej cywilizacji), ale także ideę skuteczności działania rozumów „bezrozumnych”. Samoświadomość wydaje się zbędnym ewolucyjnym naddatkiem u mechanicznych istot z powieści Lema. Nie jest im potrzebna do przetrwania, nie decyduje o ich skuteczności poczynań. Współczesne algorytmy sztucznej inteligencji również nie są świadome swego istnienia, a działają w środowisku cyfrowym o wiele sprawniej od rozumów biologicznych. Wydaje się, w rysującej się wcale nieodległej perspektywie, że najnowsze algorytmy będą w stanie obejść słynny test Turinga (o którym szerzej napiszę przy okazji rozważań nt. twórczości P. K. Dicka), pomimo tego, że nie okażą się samoświadomymi istotami (a przecież to właśnie rzeczony test miał wykazać); że będą potrafiły swobodnie kontekstowo tłumaczyć teksty z obcych języków, nie rozumiejąc ich znaczenia, że będą w stanie pisać teksty o charakterze artystycznym, nie odczuwając przy tym żadnej satysfakcji z powodu ich mniemanej doskonałości formalnej. Ilustruje ten paradoks eksperyment „chińskiego pokoju” Johna Searle’a, w którym opisany jest proces tłumaczenia na język chiński bez konieczności choćby podstawowej znajomości tegoż języka, czy też w ogóle bez świadomości, że uczestniczy się w procedurach translacyjnych (szerzej patrz: Penrose 2021: 57–58).

doświadczeń z okresu wojennego) pisarz niejednokrotnie dawał wyraz niezadowoleniu z powodu biologicznego i intelektualnego wyposażenia człowieka. Natomiast w wymiarze „mikro” ewolucja dokonała wielkich odkryć – zachwycał się Lem informacyjnym zasobem komórek organicznych, zazdrościł roślinom ich technologii absorpcji energii słonecznej, którym ludzkie rozwiązania wciąż dorównać nie mogą; także owdzia organizacja społeczna, zaprogramowane społeczeństwa mrówek, termitów, pszczoł wydawały mu się niezwykle wydolne dostosowawczo.

Występuje w *Niezwykczonym* jeszcze jeden charakterystyczny dla twórczości Lema motyw – nieprzystawalności epistemologicznej obcych sobie rozumów (choć w przypadku kolektywnej „istoty” z Regis III trudno mówić o rozumie w klasycznym tego słowa znaczeniu). Nieвозмоzne jest porozumienie na podstawowym choćby poziomie między intelektami, które ewoluowały w skrajnie odmiennych warunkach, a które wyposażone są prawdopodobnie w zupełnie różne sensorium. W powieści polskiego pisarza problem jest jeszcze głębszy – ponieważ Ziemianie napotykaają zmutowany odprysk obcej technologii, co zwielokrotnia trudności z nawiązaniem kontaktu. Skoro zaś zrozumienie nie jest możliwe, to może nie ma również konfliktu w klasycznym rozumieniu tego terminu, zakładającego przecież istnienie przynajmniej dwu zantagonizowanych stron, świadomych kolizji interesów. Uczestnicy wyprawy uświadamiają sobie, że postawa eksploratorów kosmosu nie powinna polegać na tym, by „szukać tylko podobnych sobie i takich tylko rozumieć, ale na tym, żeby nie wtrącać się do nie swoich, nie ludzkich spraw. Zagarniać pustkę, owszem, dlaczegóż by nie, ale nie atakować tego, co istnieje, co w ciągu milionoleci wytworzyło swoją własną, niepodległą nikomu ani niczemu oprócz sił promienistych i sił materialnych równowagę trwania, czynnego, aktywnego trwania, które jest ani lepsze, ani gorsze od trwania białkowych związków, zwanych zwierzętami czy ludźmi” (Lem 2008: 125). Fredric Jameson interpretował ten motyw nieprzystawalności intelektualnej w duchu Heideggera – zostawmy obcą istotę swojemu losowi, żeby „był w swoim bycie” – jako „istotę zamkniętą i nieprzeniknioną niczym przyszłość” (Jameson 2011: 140). Czyż nie takimi właśnie „nieprzeniknionymi istotami” okażą się w ostatecznym rezultacie silne sztuczne inteligencje?

Dalsze konsekwencje mikroautomatyzacji, tym razem społeczne, opisuje Lem w jednej ze swych ostatnich powieści – *Wizji lokalnej*. Akcja

utworu rozgrywa się na obcej planecie podzielonej na zantagonizowane obozy polityczne: Kurdlandię i Luzanię. W Kurdlandii funkcjonuje niezwykle rozbudowany aparat biurokratyczno-policyjny, swobody obywatelskie zaś są mocno ograniczone. Luzania z kolei jawi się na tym niekorzystnym tle istnym konsumpcyjnym rajem – oczywiście tylko na pierwszy rzut oka. Funkcjonuje tu bowiem system społeczny oparty w znacznej mierze na tzw. etykosferze, czyli etyce stosowanej empirycznie przez „bystry” – zminiaturyzowane do rozmiarów atomów automaty. W praktyce owa „nano-etyka” wygląda tak, że jakiegokolwiek czyny moralnie naganne okazują się niemożnością, ponieważ rzeczony bystry natychmiast je udaremniają (np. narzędzia potencjalnych zbrodni i niegodziwości rozsypują się na skutek atomowych interwencji w proch). Systemu luzańskiego, pomimo widocznych gołym okiem przewag nad kurdlańskim, idealnym w żadnym razie nazwać nie można, ponieważ „stwarza powszechny raj, w którym każdy siedzi z własnym piekłem w sobie i nie może dać drugiemu do posmakowania tego piekła” (Lem 1982: 282)⁸.

Ów swoisty, pojmowany zarazem bardzo dosłownie, minimalizm etyczny zbliża stanowisko Lema do poglądów Karla Edmunda Poppera (którego notabene w licznych wywiadach wymienia jako swojego „ulubionego filozofa”). To Popper w swym dziele *Spółeczeństwo otwarte i jego wrogowie* (wśród tytułowych wrogów filozof sytuował nie byle kogo, gdyż m.in. Platona i Marksa) sformułował bliskie Lemowi zasady tzw. społecznej inżynierii cząstkowej (por. Popper 1993: 181–183), którą przeciwstawia krytykowanej przez siebie inżynierii utopijnej. Popperowski inżynier cząstkowy może wprawdzie wierzyć w to, że idealne systemy społeczne zostaną wprowadzone kiedyś w życie, ale musi pamiętać, że o wiele łatwiej jest empirycznie stwierdzić istniejące zło, niż doprowadzić do powszechnej zgody w kwestii powszechnie obowiązującego dobra. Popper, podobnie jak Lem (choć może należałoby odwrócić kolejność tej zależności), nie ufa rozwiązaniom globalnym, opierającym się na prostych recepturach ogólnego błogostanu społecznego, nie ma uniwersal-

⁸ Nietrudno dostrzec analogie między sytuacją zarysowaną w powieści a światową rzeczywistością społeczno-polityczną lat 80. XX w. Kurdlandia, siermiężna i autorytarna rządzona, to przetworzony literacko obraz Związku Radzieckiego wraz z „satelitami”, do których zaliczała się Polska; Luzania zaś, ze swym dobrobytem i dosytem totalnym oraz konsumpcyjnym „rozpasaniem”, to zmetaforyzowany obraz Zachodu. Lem żadnego systemu nie faworyzuje, w obu dostrzega złe strony, choć odmiennej natury.

nego panaceum na wszystkie socjalne bolączki i choroby; tylko pewien rodzaj „pracy u podstaw” może stopniowo polepszyć jakość świata.

Zarazem jednak nie byłby zapewne Lem sobą, gdyby nawet z Popperem w swej powieści mimo wszystko nie polemizował. Czyż nie jest bowiem etykosfera narzędziem wywiedzionym bezpośrednio właśnie z zasad inżynierii cząstkowej? Czyż nie zwalczają bystry doraźnego zła (arbitralnie pojmowanego) z tak obezwładniającą doskonałością i skutecznością, że stają się jakby działającym systemem społecznym, przeobrażając się z „pogotowia aksjologicznego” w narzędzie holistyczne, całościowe, wszechogarniające? Tym bardziej, że pojawia się w powieści sugestia dalszego rozwoju etykosfery, działającej już nie tylko jako urządzenie profilaktyczne, ale również badające skrycie ludzkie profile charakterologiczne po to, by każdemu dobierać taki los, jaki najbardziej mu będzie odpowiadać⁹. Wydaje się, że ostrzega w swej powieści pisarz przed niebezpieczeństwem coraz częściej podnoszonym przez współczesnych komentatorów zmian cywilizacyjnych i nowych technologii. Nawet aksjologiczne bezpieczniki, postulowane przez przywoływanego tu Asimova, instalowane w oprogramowaniu sztucznych inteligencji, mogą nie dawać gwarancji bezpieczeństwa rodzajowi ludzkiemu. Mało tego – zbyt ogólnie (lub zbyt szczegółowo) formułowane – mogą w dalszych, niemożliwych do przewidzenia konsekwencjach być przyczyną niewyobrażalnych katastrof.

Wydaje się, że naturalną konsekwencją syntetycznej ewolucji, jakby ostatecznym jej rezultatem, jest stan, w którym cała otaczająca nas rzeczywistość jest tworem sztucznym, rozumnie przekonstruowanym. Lem opisuje taki stan rzeczy w *Nowej kosmogonii*, utworze należącym do serii jego opowieści apokryficznych opublikowanych w tomie *Doskonała próżnia*. Tekst jest pomyślany jako przemówienie odbierającego nagrodę Nobla wybitnego astronoma Alfreda Testy, który opowiada o bardzo niepokojącej wizji kosmosu. W świetle jego hipotezy wszechświat – wraz ze wszystkimi swymi nie do końca poznanymi przez ludzkość fenomenami fizycznymi (jak czarne dziury, kwazary, pulsary itp.) – okazuje się w znacznej swej części tworem sztucznym, jego kształt jest

⁹ Koresponduje to nieodległe i proroczo ze współczesną nam epoką Big Data, algorytmami gromadzącymi w sieci niezliczone informacje na temat ludzi i ich preferencji. W USA weszły już w użycie tzw. algorytmy sądowe, które mają za zadanie brać pod uwagę nie tylko popełnione przez oskarżonych realne złe uczynki, ale także ekstrapolować przyszłe – i do owych ekstrapolacji dostosowywać wymiar kary.

wynikiem swoistego „kompromisu” niezwykle technologicznie rozwiniętych cywilizacji. Supercywilizacje, działając we własnym dobrze pojętym interesie, miały wytworzyć, w monstualnym i powolnym procesie ewolucyjnych (acz sterowanych odgórnie) zmian, nieistniejący wcześniej porządek i prawa, które dziś ludzkość odkrywa i uważa – błędnie – za naturalne, wieczne, a nawet majestatyczne prawa natury. Dawniej rządzić miały światem zupełnie inne prawa fizyczne, różniące się między sobą w zależności od strefy kosmosu. Przykładem ujednoczenia może być stała wartość prędkości światła, która u zarania dziejów wszechświata bynajmniej nie stanowiła powszechnie nieprzekraczalnej granicy (jako najwyższa możliwa do osiągnięcia prędkość). Tymczasem starsze od ludzkości o miliony, jeśli nie miliardy lat supercywilizacje, zdobywszy wiedzę naukową i technologiczną niewyobrażalną z ludzkiego punktu widzenia, odkryły wzajem swoje istnienie i rozpoczęły po dziś dzień trwającą grę, do której dołącza się, chcąc nie chcąc, każda cywilizacja dostatecznie dojrzała, by wziąć w niej udział. Na przestrzeni miliardów lat – po spowodowaniu niewyobrażalnych katastrof (uboczne skutki monstualnych eksperymentów) – wykreowano wreszcie jednolity kosmos, w którym prawa fizyki są wszędzie jednakowe. Kosmiczne *status quo* wymaga, by wszechświat był rządzony surowymi regułami czasoprzestrzeni, wraz z barierą prędkości światła, która z kolei sprawia, że komunikacja – po przekroczeniu określonego dystansu – staje się praktycznie niemożliwa. Być może uczestnicy owej gargantuicznej gry dali milczące przyzwolenie na wykopanie między sobą owej komunikacyjnej przepaści, co przy okazji tłumaczy tajemnicę *silentium universi*. Z tych powodów w zasadzie nierozróżnialne jest we „współczesnym” wszechświecie to co „naturalnym”, a co „sztucznym” się jawi.

Poza tym wszystkim proces gry pobudza bezwiednie negentropię, a to sprzyja eksplozjom życia w różnych częściach kosmosu; może się zatem zdarzyć, że ryzykowna gra doprowadzi wreszcie do zamknięcia „rozgrywki”, czyli końca świata, jeśli np. przyłączą się do niej przedwcześnie cywilizacje nie dość dojrzałe, by respektować zastany *status quo*. Gdy już wszystko powróci do stanu początkowego chaosu, po eonach być może pojawi się kolejna grupa wyrafinowanych graczy – i procesy porządkowania zaczną się od nowa. W takim razie w ogóle nie może być mowy o rzeczywistym początku wszechświata – wszystko płynie, wszystko się powtarza niczym w systemie ontologicznym Heraklita,

ale porządkującą rolę *logosu* odgrywają w tym akurat wypadku rzeczowe supercywilizacje.

Idee Lema wiodące od stopniowej miniaturyzacji automatów, które stają się jakby aktywną osnową materii, aż po konstruktorskie racjonalizowanie całego kosmosu przywodzą na myśl koncepcje starożytnych stoików, ich dynamiczne pojmowanie materii (*pneumy*), która działa celowo. *Logos* jest tu niejako „rozpuszczony” w przyrodzie, racjonalność okazuje się zaś zarówno „oddolną” zasadą funkcjonowania rzeczywistości, jak i jej finalnym celem. To panteistyczne pojmowanie wszechświata wydaje się rzeczywiście bliskie ideom Lema, a w każdym razie dalszym mniemanym konsekwencjom, jakie z nich wynikają.

Jest to niepokojąca i silnie działająca na wyobraźnię wizja rzeczywistości, nie tylko z tego powodu, że ukazuje możliwość periodycznego i hierarchicznego funkcjonowania świadomości w kosmosie (piętrowość rozumów może ciągnąć się wszak w nieskończoność), ale również dlatego, że istnieje tu domniemanie niemożności rozróżnienia tego, co sztuczne i wykreowane, od tego, co przypadkowe i naturalne. Wizja wszechświata racjonalnego, rzeczywistości „skonstruowanej” celowo, znajduje swoje odbicie u jednego z najbardziej znanych przedstawicieli współczesnego transhumanizmu, Raya Kurzweila, który kreśli w swej słynnej książce *Nadchodzi Osobliwość* (rzeczona „osobliwość”, pisana zawsze wielką literą, rozumiana tam jest jako rekurencyjne samodoskonalenie sztucznych inteligencji prowadzące szybko do zaistnienia fenomenu syntetycznej superinteligencji, do czego ma dojść w przewidywalnych horyzontach czasowych) swoistą historiozofię, periodyzując epoki w rozwoju wszechświata. Otóż szósta, dodajmy, że ostatnia, epoka ma nosić umowną, lecz znaczącą nazwę „Wszechświat się budzi”. Tak pisze o tym epokowym momencie sam Kurzweil:

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 poprzez reorganizację materii i energii w celu zapewnienia najlepszego poziomu przetwarzania [...] po to, by rozprze-strzeniać się ze swojego miejsca pochodzenia na Ziemi.

Obecnie rozumiemy prędkość światła jako czynnik ograniczający transfer informacji. Obejście tego ograniczenia jest dziś uważane za wysoce wątpliwe, są jednak wskazówki sugerujące, że to ograniczenie można wyeliminować. Jeśli pojawią się choćby niewielkie odchylenia,

ostatecznie wykorzystamy tę nadświatłą zdolność. To, czy nasza cywilizacja będzie nasycać resztę wszechświata swoją kreatywnością i inteligencją szybciej czy wolniej, zależy od jej niezmienności. W każdym razie „głupia” materia i mechanizmy wszechświata będą przekształcone w znakomicie wysublimowane formy inteligencji, które ustanowią szóstą epokę ewolucji wzorców informacji. Jest to ostateczne przeznaczenie Osobliwości i wszechświata (Kurzweil 2013: 35).

Philip K. Dick: sztuczna inteligencja jako funkcja szaleństw

Znaczenie pisarstwa Dicka daleko wykracza poza stereotypowe osiągnięcia literatury pulpowej. Jest to twórczość przesiąknięta tematyką filozoficzną, a przede wszystkim – jak słusznie dostrzega to polski pisarz i komentator twórczości Dicka Marek Oramus – dylematami z zakresu teorii poznania, gnoseologii.

Różnica między naszym statecznym bytowaniem a szarpaniną nie-szczęśników Dicka polega na tym, że u nas uprawianie filozofii należy jakby do życiowych luksusów, gnoseologia jest czymś, bez czego można się doskonale obejść, natomiast wewnątrz perfidnych światów Dicka umiejętne jej uprawianie staje się kwestią przeżycia (np. w *Ubiku*). [...] W świecie Dicka filozofia schodzi niejako na ulicę i staje się dla każdego zjadacza chleba sprawą nie mniej palącą, niż dla nas jest nią na przykład zagrożenie biosfery (Oramus 1987: 53).

Dick był twórcą kwestionującym realność otaczającej go rzeczywistości, a przy tym zadającym niepokojące pytania na temat tożsamości człowieka. Domeną jego twórczości jest rozpad rzeczywistości, będący jakby projekcją stanów wewnętrznych bohaterów jego opowieści (i to bez względu na to, czy mamy do czynienia z kreacją literacką istoty ludzkiej, czy sztucznej inteligencji – komputera, androida, robota); mało tego, owe projekcje doznają w prozie Dicka ontycznego wzmocnienia – stają się zasadą porządkującą bądź – częściej – dezorganizującą świat przedstawiony. Innymi słowy, powieściowe stany rzeczywistości fabularnych Dicka przypominają obrazy psychicznych meandrów człowieka chorego¹⁰; w jakiejś mierze można je porównać do strategii twórczych

¹⁰ Stany chorobowe amerykańskiego pisarza są dość dobrze udokumentowane (był podejrzewany o schizofrenię, chorobę Aspergera, niewątpliwie zażywał mnóstwo nar-

surrealistów. Lem porównuje światy autora *Trzech stygmatów Palmera Eldritch* do rozdartego i targanego wewnętrznymi sprzecznościami umysłu schizofrenika. We wczesnym opowiadaniu amerykańskiego pisarza zatytułowanym *Nul-0* (1958) pojawia się opis osobowości paranoicznej, ale nie w pejoratywnym tego słowa znaczeniu – w świecie tej fabuły paranoja zapewnia życiowy sukces. Dick tak opisuje ten typ osobowości:

Niewiarygodne. Sam z trudem mogę w to uwierzyć. Twoje działania są na wskroś przesiąknięte logiką. Nie dopuszczasz do siebie żadnych emocji. Twój umysł doszczętnie pozbawiony jest wszelkich moralnych bądź kulturowych uprzedzeń. Jesteś doskonałą jednostką paranoidalną, niezdolną do empatii. Cechuje cię całkowity brak możliwości odczuwania smutku, żalu, współczucia czy jakichkolwiek emocji zazwyczaj cechujących człowieka. [...] To przechodzi ludzkie wyobrażenie. Twoja superlogika funkcjonuje z dala od stronniczego nastawienia będącego na ogół zasadniczym kryterium oceny. Uważasz, że cały świat sprzyścił się przeciwko tobie. [...] Zawsze klasyfikowano paranoję jako chorobę umysłową. Ale przecież nią nie jest! Nie następuje załamanie kontaktu z rzeczywistością – wręcz przeciwnie, paranoik jest bezpośrednio z nią związany. To doskonały empirysta. Z umysłem niezasmieconym naleciałościami natury etycznej bądź moralno-kulturowej. Paranoik dostrzega rzeczy w ich rzeczywistej postaci; to w istocie jedyny człowiek przy zdrowych zmysłach (Dick 2002: 163).

Jak się wydaje, do pewnego stopnia fragment ów oddaje także nastawienie Dicka do fenomenu sztucznej inteligencji; w jakimś sensie ilustruje również powszechne wyobrażenie na temat „wizerunku” sztucznych rozumów. Mają to być intelekty wyzbyte sfery emocjonalnej, funkcjonujące w sposób wysoce pragmatyczny, opierające się w swych wnioskowaniach na empirycznych dowodach, a nie na przekonaniach. Przypominają nieco ideał stoickiego mędrca – wyzbytego afektów, nieosiągalnego dla codziennych trosk, oddalonego od życia, ale czy przypadkiem nie oddalonego także od... człowieczeństwa? Podobny status w tekstach amerykańskiego pisarza usiłują osiągać rów-

kotyków, w tej liczbie amfetaminę i LSD). Sporo miejsca poświęca tym zagadnieniom biograf Dicka, Anthony Peake (por. Peake 2015: 94, 112–113, 298, 322). Fascynujące u Dicka jest to, że potrafi przetworzyć swoje halucynacje, stany chorobowe na bardzo oryginalną twórczość. Nie jest oczywiście przypadkiem odosobnionym – wystarczy wspomnieć Maupassanta, Dostojewskiego, van Gogha i wielu innych.

niez ludzie – jest w tym pewien paradoks, im bardziej bowiem staramy się osiągnąć ideał człowieczeństwa „ulepszonych”, tym bardziej oddalamy się od bycia człowiekiem. Teksty Dicka można oczywiście interpretować również jako alegorie – jest w twórczości autora *Ubika* zawarty, wspomniany już na łamach tego tekstu, efekt antropologicznego zwierciadła – ludzie przeglądają się w lustrze kreacji sztucznych inteligencji, dostrzegając uwypuklone cechy negatywne własnej natury. W końcu komputery i roboty są przedłużeniem nas samych, lecz również naszym paradoksalnym zaprzeczeniem i – być może – kresem. Jak słusznie zauważa Magdalena Radkowska-Walkowicz, kreacje robotów w pewnym (wczesnym) segmencie twórczości Dicka odzwierciedlają „bardzo popularne spojrzenie na sztucznych ludzi: są bez duszy, bez moralności, bez uczuć. Istoty takie obwieszczają niechybny koniec człowieczeństwa” (Radkowska-Walkowicz 2008: 157).

Tematykę sztucznej inteligencji podjął Dick w swej bogato reprezentowanej nowelistyce, a ukoronowaniem tych rozważań była powieść *Czy androidy śnią o elektrycznych owcach?* (1968), na podstawie której powstał głośny film w reżyserii Ridleya Scotta *Blade Runner* (1982). Pisarz niejednokrotnie przestrzegał przed alienacją syntetycznej rzeczywistości. W swoistym posłowniu do opowiadania, którego nie będziemy tu akurat szerzej omawiać (*Kolonia*), pisał: „Maksymalna paranoja nie polega na przekonaniu, że wszyscy są przeciwko tobie. Zamiast mówić »Mój szef spiskuje przeciwko mnie«, mówisz wtedy: »Telefon mojego szefa spiskuje przeciwko mnie«. Przedmioty i tak niekiedy zdają się posiadać własną wolę. [...] W tym opowiadaniu starałem się stworzyć sytuację, która w racjonalny sposób wyjaśniłaby zagadkę spisku przedmiotów przeciwko ludziom, bez odwoływania się do ułomności ludzkiej psychiki” (Dick 1998: 461). To obsesyjne poczucie wyobcowania pośród świata rzeczy owocować musiało w twórczości Dicka pesymistycznymi scenariuszami wpływu sztucznych inteligencji (czyli przedmiotów upodmiotowionych, obdarzonych świadomością) na świat ludzi. W opowiadaniu *Obrońcy* pisarz przedstawia świat po konflikcie nuklearnym między Związkiem Radzieckim a Stanami Zjednoczonymi. Mieszkańcy miast zmuszeni zostali do ukrycia się w podziemnych schronach, gdzie wiodą ascetyczną egzystencję. Na powierzchni pozostały tzw. ołowiaki, czyli autonomiczne maszyny skonstruowane przez ludzi w celu prowadzenia działań wojennych. Przez wiele lat tylko one wy-

chodziły na powierzchnię i relacjonowały swym programistom rezultat wojny. Okazuje się jednak, że od wielu dziesięcioleci nie ma żadnego konfliktu, a relacje maszyn są fałszywe. Sztuczne istoty od dawna nie prowadzą działań wojennych – od momentu, gdy ludzie schronili się pod powierzchnią planety, zapanował pokój. Ołowiaki dokonały pragmatycznej analizy fenomenu wojny, dochodząc do jedynego możliwego wniosku, że jest ona pozbawiona jakiegokolwiek sensu. Roboty zbadały również naturę człowieka w kontekście historycznym – rezultaty nie okazują się dla gatunku *Homo sapiens* korzystne – ludzie zawsze byli istotami agresywnymi, z silnie rozwiniętym instynktem terytorialnym. Postanowiono więc odizolować od miast jednostki, które przetrwały konflikt, i zaprowadzić ład. Ład ów ma obejmować nie tylko wymiar materialny rzeczywistości, ale również społeczny. Zjednoczenie ludzkości ma być bowiem, w interpretacji syntetycznych rozumów, ostatecznym celem historii; i zostanie dokonane bez względu na opory ludzkości. Wymowa tekstu jest dwuznaczna – sztuczne inteligencje pętają wprawdzie ludzkość, ale „dla jej dobra”; motyw ten będzie wielokrotnie powtarzany w różnych konfiguracjach w światowej fantastyce (wystarczy wspomnieć – omawianą tu już – *Wizję lokalną* Lema czy głośny film Wachowskich *Matrix*).

W nieco innym kontekście i innej roli obsadzona jest syntetyczna inteligencja w opowiadaniu *Święty spór*. „Bohaterem” tej fabuły jest samoświadomy komputer Genux-B, zaprojektowany jako obrońca ludzkości przed możliwymi zagrożeniami zewnętrznymi. Odkrywa on diaboliczny spisek związany z osobą niejakiego Herberta Sousy, producenta automatów rozprowadzających gumę do żucia. Wywiad wojskowy w żaden sposób nie jest w stanie potwierdzić apokaliptycznych przewidywań maszyny myślącej – Sousa nie przejawia żadnych agresywnych zachowań, gumy sprzedawane w sieci jego automatów nie posiadają niepokojących składników chemicznych. Programiści dochodzą do wniosku, że Genux-B popadł w paranoję, co jest o tyle niebezpieczne, że komputer stanowi istotne ogniwo systemu bezpieczeństwa Stanów Zjednoczonych. Postawiona zostaje hipoteza, że obłąd komputera może być wynikiem działań dywersyjnych – fikcyjny spisek ma zaabsorbować jego uwagę w taki sposób, by nie dostrzegł realnego niebezpieczeństwa – np. ataku bombowego na terytorium USA. System obronny jest uzależniony od sprawnego funkcjonowania Genuxa-B, niełatwo

jest zatem maszynę po prostu wyłączyć; programiści podejmują zatem brawurową próbę resetu jej odbioru rzeczywistości. Preparują informację, jakoby Herbert Sousa nigdy nie istniał – komputer uznaje jednak, że jest to informacja fałszywa. Kolejnym krokiem jest próba przekonania komputera, że on sam nie istnieje; otrzymują w odpowiedzi wykład ontologiczny, którego nie powstydziliby się sam Kartezjusz:

ANALIZA INFORMACJI NA TEMAT NIEISTNIENIA WIELOCZYNNIKOWYCH INSTRUMENTÓW ANALITYCZNYCH TYPU GENUX-B. JEŻELI INFORMACJA NUMER 340s70 JEST PRAWDZIWA, TO: JA NIE ISTNIEJE. JEŻELI JA NIE ISTNIEJE, TO NIE MA SPOSOBU, ABY POINFORMOWAĆ MNIE O NIEISTNIENIU MOJEJ KLASY RODZAJOWEJ. JEŻELI NIE MOŻNA MNIE O TYM POINFORMOWAĆ, TO NIE POINFORMOWALIŚCIE MNIE I Z MOJEGO PUNKTU WIDZENIA INFORMACJA NUMER 340s70 NIE MA RACJI BYTU. DLATEGO: JA ISTNIEJE (Dick 2000: 167)¹¹.

Problem związany z możliwościami percepcyjnymi i analitycznymi Genuxa-B jest szerszy – posiada on bowiem dostęp do niezliczonej ilości danych, które z ludzkiego punktu widzenia wydają się nieistotne, natomiast z punktu widzenia sztucznej inteligencji nabierają korelatywnej istotności. Nadmiar impulsów środowiskowych może być przyczyną szaleństwa maszyny, i taka jest hipoteza programistów, ale może również okazać się przyczyną poszerzonej perspektywy. Dick intuicyjnie przewiduje tu efekty związane z fenomenem *big data* i rewolucją w sposobie postrzegania (i badania) rzeczywistości. Mniej istotne stają się we współczesnych procesach odbioru rzeczywistości, jak opisują to Viktor Mayer-Schonberger i Kenneth Cukier, związki przyczynowo-skutkowe, do których ludzki umysł tak bardzo – na mocy przystosowań ewolucyjnych – jest przyzwyczajony, znaczenia natomiast nabierają, przypadkowe na pierwszy rzut oka, pozornie chaotyczne korelaty zjawisk, które można zbadać, wyłącznie dysponując potężną mocą obliczeniową (por. Mayer-Schonberger, Cukier 2014: 53–55).

¹¹ Podobny motyw występuje w filmie *Ciemna gwiazda* (1974) w reżyserii Johna Carpentera – bohater usiłuje powstrzymać eksplozję inteligentnej bomby atomowej, wmawiając jej, że ta nie istnieje. W dyskusji pada argumentacja filozoficzna – poglądy Kartezjusza, Husserla, omawiane są paradoksy solipsyzmu. Oczywiście jest to groteska, która ilustruje jednak „uwikłanie” tematyki sztucznej inteligencji w konteksty ontologiczno-epistemologiczne.

Motyw komputerowego szaleństwa¹² rozwija autor *Człowieka z Wysokiego Zamku* w groteskowym opowiadaniu zatytułowanym *Dzień, kiedy pan komputer stracił rozum*. Przedstawiona jest tu rzeczywistość postapokaliptyczna, będąca wynikiem konfliktu nuklearnego (który w rzeczywistości fabularnej miał miejsce już na początku lat 80. XX wieku); zarazem przesycona elektroniką i zarządzana przez scentralizowany system sztucznej inteligencji.

Kontrolowanie wszystkich urządzeń z centralnego źródła od początku, czyli od 1982 roku, było kiepskim pomysłem. Rzecz jasna, zasadnicza koncepcja brzmiała niezłe: przy znikomej ilości warstwy ozonowej zbyt wielu ludzi wykazywało tendencję do irracjonalnych zachowań, dlatego konieczne stało się rozwiązanie problemu za pomocą elektronicznych środków odpornych na mordercze promieniowanie ultrafioletowe, które zalewało Ziemię. Wtedy wydawało się, że pan Komputer to ostatnia deska ratunku. Niemniej, przykro to mówić, pan Komputer przejął zbyt wiele szkodliwych cech od swoich konstruktorów i – podobnie jak oni – również jego nachodziły okresy psychicznej niedyspozycji (Dick 2000: 318).

W opowiadaniu *Ucieczka* można zaobserwować przynajmniej dwie charakteryzujące twórczość Dicka obsesje – po pierwsze, krytykę kapitalistycznego konsumpcjonizmu, która przewija się niemal przez całą

¹² Chyba najciekawszą ilustracją opisu maszynowego szaleństwa jest kreacja HALa 9000, fikcyjnego komputera przedstawionego w książce Arthura C. Clarke'a *2001: Odyseja kosmiczna* oraz w słynniejszym jeszcze filmie w reżyserii Stanleya Kubricka o tym samym tytule. W powieściowej fabule zostaje uruchomiony w laboratoriach HAL w Urbana, Illinois, w roku 1997 (znacznie wyprzedzając, jak się okazało, rzeczywiste osiągnięcia na polu konstrukcji sztucznej inteligencji). Stanowić ma największe osiągnięcie przemysłu informatycznego i najbardziej zaawansowany projekt sztucznej inteligencji. Zostaje zainstalowany na statku kosmicznym Discovery One jako kontroler misji, zarządzający wszystkimi aspektami technicznymi wyprawy – jako jedyny zna prawdziwy jej cel (którym jest zbadanie śladów zaawansowanej obcej cywilizacji). Jego interfejs został zoptymalizowany pod kątem komunikacji z człowiekiem – posiada silnie rozwinięty podsystem emocjonalny. HAL, z nie do końca znanych przyczyn (w późniejszych rozwinięciach fabuły okazuje się, że jego „szaleństwo” było wynikiem otrzymania z ziemskiej centrali sprzecznych poleceń, co sprawiło niemożność ich uzgodnienia – HAL popadł w tzw. pętlę Hofstadtera-Möbiusa), buntuje się i zabija wszystkich ludzkich członków załogi, z wyjątkiem dowódcy, Davida Bowmana, który z kolei „morduje” HALa, wyłączając wszystkie jego wyższe funkcje. W głośnym filmie Kubricka niezwykle poruszająca jest scena „agonii” HALa – jego strach przed utratą „życia”, powolne staczenie się wybitnej inteligencji na coraz niższe szczeble świadomości, demencji. Jest w tej scenie coś niesłychanie ludzkiego.

twórczość pisarza. Po drugie zaś wizerunek wyobcowanej sztucznej inteligencji. Jest to świat wszechobecnej reklamy. „Handlujące roboty. [...] Roboty i reklamy wizualne drażące człowiekowi mózg. Snują się za ludźmi jak cienie, dopóki nie wyzioną ducha” (Dick 2002: 208). Maszyny są niestrudzone w realizacji wytycznych swego – bywa, że absurdalnego – oprogramowania. Bohater opowiadania, Ed Morris, pragnie uciec od wszechobecnych reklam, a realizuje swój cel w sposób radykalny – planuje mianowicie przenieść się poza obręb Układu Słonecznego i osiedlić, wraz z żoną, na odległej kolonii. W jego mieszkaniu pojawia się jednak robot, który zachwala nowy tajemniczy produkt o nazwie „fasrad”. Narasta poczucie absurdu – nie wiadomo, czym rzeczony fasrad jest w istocie, wiadomo natomiast, że jest „absolutnie niezbędny i niezastąpiony”, a przy tym uniwersalny (przyda się rodzinie nawet w przypadku eksplozji nuklearnej). Fasrad może również zastępować swego właściciela w miejscach publicznych – np. pełnić jego obowiązki w pracy przez dziesięć dni. Oczywiście w końcu okazuje się, że rzeczonym fasradem jest „samosprzedający się” robot (Funkcjonalny Automatyczny Samoregulujący Android). Odmowa kupna nie wchodzi w rachubę, dopóki klienci go nie kupią, dopóty nie mogą wydawać mu poleceń – wytwarza się błędne koło wzajemnych zależności. Fasrad odpowiada przed Korporacją Samoregulujących się Androidów, organizacja ta tak zaprogramowała robota, by ten, bez względu na okoliczności, został u potencjalnych nabywców do czasu, aż zostanie kupiony – potem okres wydłuża się w zależności od umowy. Nie ma odwrotu od nowego wspaniałego świata wypełnionego niestrudzonymi fasradami. Pewnym – wątpliwym – pocieszeniem jest fakt, że mimo swej uniwersalności androidy nie są w stanie odpowiedzieć na wszystkie możliwe pytania; niektóre problemy pozostają poza ich zasięgiem pojmowania. „Czasem ktoś potrzebuje natychmiastowej odpowiedzi na swoje pytanie; fasrad zawsze gotów jest rozwiązać problem natury teoretycznej bądź praktycznej. Wszystko oprócz kwestii metafizycznych” (Dick 2000: 214). W późniejszych swych tekstach Dick zniesie także i te ograniczenia.

Tak właśnie jest w noweli zatytułowanej *Elektryczna mrówka*. Jej bohater, Garson Pool, uchodzący przez całe swe dość bezbarwne życie za zwyczajnego człowieka, nieoczekiwanie odkrywa, że bynajmniej człowiekiem nie jest. Jest natomiast tytułową „elektryczną mrówką”, czyli androidem do złudzenia przypominającym istotę ludzką, którego

dłoń, gdyby się jej baczniej przyjrzeć, okazuje się „z pozoru organiczna, pokryta warstwą naturalnej skóry i wypełniona krwią pulsującą w żyłach, w istocie kryła w sobie sieć miniaturowych elementów, druty i obwody” (Dick 2000: 230). Odkrywszy swą sztuczność, bohater przeprowadza na sobie radykalne eksperymenty – otwiera czaszkę i odkrywa, że jego doznania zapisane są na perforowanej „taśmie rzeczywistości”, którą nakłuwa, usuwa jej fragmenty, obserwując fenomenalne zmiany w otaczającym go świecie. Ma ogromne ambicje poznawcze, jego głód wiedzy bardzo przypomina podobne ludzkie pragnienia. „Mam przed sobą możliwość doświadczenia wszystkiego. Jednocześnie. Poznać wszechświat w całości, zetknąć się z całą rzeczywistością. To coś, czego nie dokonał żaden człowiek. Koncert symfoniczny dobiegający spoza czasu, wszystkie nuty i instrumenty rozbrzmiewające jednocześnie. [...] Chciałbym przejąć kontrolę nad czasem” (Dick 2000: 240).

Eksperymentując z obrazem rzeczywistości, bohater noweli Dicka doprowadza się do autodestrukcji – żądza poznania prawdy ponad wszystko ma w sobie zalążki samozniszczenia. Pojawiają się zarazem akcenty charakterystyczne dla późniejszej twórczości amerykańskiego pisarza – okazuje się bowiem, że istnieje jeszcze zewnętrzny obserwator zmagających epistemologicznych Poola – kobieta, Sarah Benton, która tuż po autodestrukcji „elektrycznej mrówki” dostrzega z przerażeniem, że sama przestaje istnieć. Pomimo subiektywnej świadomości własnej odrębności okazuje się tylko wiązką wrażeń Poola, elementem jego zamkniętego poznawczo oprogramowania. Być może cały otaczający świat jest tylko rodzajem cyfrowej ułudy, android zaś okazuje się kimś w rodzaju kalekiego demiurga, nieświadomego swej zadziwiającej kondycji.

Kwestia tożsamości sztucznych inteligencji stanowi oczywiście główny temat jednej z najsłynniejszych powieści Dicka *Czy androidy śnią o elektrycznych owcach?*, ale warto może jeszcze kontekstowo wspomnieć o innym utworze amerykańskiego pisarza – mianowicie *Możemy cię zbudować*. Książka, napisana w roku 1962, a opublikowana równo dekadę później, stanowi, w opinii Jamesona, niejako wstęp do rozważań zawartych w *Czy androidy...* (Jameson 2011: 325). Jest to opowieść o firmie handlującej elektronicznymi klawikordami, której właściciele nagle postanawiają poszerzyć swój asortyment, produkując doskonale kopie postaci historycznych, mających stanowić główną atrakcję swobodnego historycznego „parku rozrywki” (Jameson zauważa tu istotną

zbieżność – powieść Dicka została opublikowana zaledwie kilka lat po otwarciu Disneylandu w roku 1955). Pierwszymi homunkulusami powołanymi do istnienia są: prezydent Lincoln i jego doradca Stanton. Maszyny mają charakter autonomiczny, mówią „własnym” głosem, jednak oczywiście (jak miało to miejsce w przypadku omawianej tu już noweli Lema *Maska*) są zarazem niewolnikami swojego oprogramowania, narzuconego przez konstruktorów. Jednak zniewolenie androidów z powieści Dicka zyskuje dodatkowy wymiar, jest jakby zwielokrotnione i dodatkowo naznaczone piętnem swoiście pojmowanego tragikomizmu – są to bowiem istoty quasi-ludzkie odtworzone na podstawie niepełnych i niedoskonałych źródeł historycznych (podstawą ich funkcjonowania, zaznacza ironicznie Dick, staje się sentymentalna czterotomowa biografia Lincolna pióra Carla Sandburga). Ich osobowość wydaje się szczątkowa, pokawałkowana, naznaczona sentymentalizmem. „Obydwie postacie charakteryzuje właściwa im bezinteresowność dająca się bez trudu wytłumaczyć ich stanem na pograniczu maszyny i trupa” (Jameson 2011: 446). Zatem pytanie o granice autonomiczności maszyn zyskuje tu kolejne piętra znaczeń – inteligentne kopie postaci historycznych, wyprodukowane z powodów komercyjnych, nie tylko nie mówią „własnym głosem”, one nawet nie mówią głosem cudzym, czyli głosem imitowanych przez nich ludzi; jest to spreparowana osobowość stereotypowego obrazu tychże postaci historycznych. Fascynujące w powieści są – skazane na porażkę – próby wydobywania się sztucznych Lincolna i Stantona z kokonu tych wielokrotnych splecionych uwarunkowań i poszukiwanie własnej tożsamości, ponieważ mimo wszystko okazują się oni istotami inteligentnymi i samoświadomymi.

Jeszcze wyraźniej te dylematy poszukiwań tożsamości wybrzmiały w powieści *Czy androidy śnią o elektrycznych owcach?* Fabuła opowiada o czasach po wojnie atomowej, która doprowadziła przede wszystkim do wymarcia zwierząt – posiadanie na własność żywego stworzenia staje się synonimem luksusu i symbolem zamożności. O wiele tańsze i powszechnie dostępne okazują się cybernetyczne imitacje istot żywych. Produkowane też są androidy do złudzenia przypominające ludzi, które wykorzystuje się – pomimo ich inteligencji i samoświadomości – do niewolniczej pracy. Co jakiś czas sztuczne istoty podnoszą bunt, uciekają, usiłując wtopić się w ludzkie otoczenie, wówczas tropią ich specyficzni łowcy – *blade runnerzy*. Jednym z takich łowców jest główny

bohater powieści, Rick Deckard¹³. Wydaje się, że w rzeczywistości powieściowej następuje zrównanie statusów maszyn i ludzi – na paradoksalnym jednak poziomie. Dick uważa, że cywilizacja ma na ludzi wpływ dehumanizujący, pozbawiający ich człowieczeństwa, cokolwiek to słowo miałooby w istocie oznaczać. W przypadku maszyn, zyskujących coraz bardziej wyrafinowaną inteligencję, proces przebiega wektorowo odwrotnie – stopniowo humanizują się. W powieści Dicka androidy wydają się czasem bardziej ludzkie niż ludzie. Stąd trudności poznawcze Deckarda, gdy ten próbuje ustalić status ściganych przez siebie zbiegów. Zresztą zarówno w powieści, jak i w jej późniejszej ekranizacji dokonanej przez Ridleya Scotta pojawiają się dość wyraźne wątpliwości dotyczące ludzkiej kondycji samego Deckarda – być może jest on androidem nieświadomym własnej sztuczności?

Najbardziej znanym sposobem sprawdzającym umowny poziom człowieczeństwa maszyn jest test Turinga, którego nazwa wywodzi się od nazwiska jego twórcy, wybitnego matematyka i pioniera informatyki. Test ów, w dużym skrócie, ma wykazać stopień nieodróżnialności ludzkiego i nieludzkiego interlokutora (szerzej zob.: Penrose 2021: 40–42). W powieści Dicka mamy do czynienia z bardziej zaawansowaną metodologią wykrywania maszyn – test Voigta-Kampfffa, który ma za zadanie stwierdzać opóźnienie w empatycznych reakcjach sztucznych istot na krzywdę zwierząt. Test „sprawdza rozszerzenie naczyń włoskowatych w tkance twarzy. Wiemy, że jest to bezwarunkowy odruch, tak zwany rumieniec, będący reakcją na bodziec, którym jest wstrząs moralny. Reakcji tej nie można kontrolować w świadomy sposób, podobnie jak przewodnictwa elektrycznego skóry, rytmu oddechu czy też pracy serca” (Dick 1999: 50). Jednak i w wypadku tej zaawansowanej metodologii rozpoznawczej istnieje możliwość błędu – okazuje się, że próba nie daje rozstrzygających rezultatów w przypadku „testowania starannie dobranej grupy schizoidalnych i schizofrenicznych pacjentów-ludzi. Szczególnie tych, którzy zdradzają objawy tak zwanego spłaszczenia

¹³ Radkowska-Walkowicz zwraca uwagę na podobieństwo brzmieniowe nazwiska głównego bohatera powieści do nazwiska Rene Descartes’a. „I nie chodzi tu tylko o podobieństwo brzmieniowe nazwisk. To właśnie Kartezjusz zafundował nam myślenie o ciele ludzkim jako o maszynie, przyczynił się do ugruntowania poglądu, że człowiek to układ mechaniczno-ciepłno-hydrauliczny” (Radkowska-Walkowicz 2008: 228). Poza tym dylematy, z jakimi przyjdzie się mierzyć Deckardowi (badanie stopnia człowieczeństwa uciekinierów, kwestie tożsamościowe), są wprost wywiedzione z dzieł francuskiego filozofa.

rezultatu” (Dick 1999: 49). Oznacza to, że istnieje grono ludzi prawie nieodróżnialnych od rozumnych maszyn.

Androidy Dicka poszukują swej ludzkiej tożsamości – chyba dobitniej jest to zaznaczone w ekranizacji powieści dokonanej przez Scotta. Androidy (w filmie nazywane replikantami) pieczołowicie przechowują pamiętki z własnej niebyłej przeszłości (ich długość istnienia jest fabrycznie zdeterminowana i wynosi cztery lata) – fotografie z nigdy niezaistniałego dzieciństwa, spreparowane wspomnienia. Ich zachowania są nacechowane specyficzną mieszanką okrucieństwa i poświęcenia, w czym zresztą bardzo przypominają ludzi. Jeden z filmowych replikantów poświęca własne istnienie, by ratować życie prześladowającego go Deckarda – są w tych scenach akcenty niemal metafizyczne (dłoń sztucznej istoty zostaje przebita gwoździem – czyżby był cierpiącym nie za swoje winy mesjaszem replikantów?). Jak ujęła to Radkowska-Walkowicz, android „wypracował sobie duszę” (Radkowska-Walkowicz 2008: 220). Dramat egzystencjalny myślących maszyn w powieści Dicka polega na ich swoistym braku „ontologicznego umocowania” w rzeczywistości. „Replikanci, roboty, androidy, homunkulusy powstają z niczego, są wyjęte z kontekstu, umieszczone poza normalnym życiem” (Radkowska-Walkowicz 2008: 221). Ich „życie” polega na nieustannej szamotaninie, poszukiwaniu sensu, zakresu swobód, niezależności od determinizmów krótkiego istnienia i oprogramowania. Znow – czyż nie są odbiciem swych stwórców, powtórzeniem tych samych dylematów umieszczonych w nieco tylko innym kontekście?

Zakończenie

Dwie wizje sztucznych inteligencji – obie skrajnie odmienne¹⁴. Wizja Lema ma charakter racjonalistyczny – sztuczna inteligencja to uciele-

¹⁴ Warto również zwrócić uwagę w tym kontekście na wizję człowieczeństwa prezentowaną przez obu pisarzy w ich twórczości. Ludzie w utworach Dicka są, jak już to zaznaczyłem, często mniej ludzcy niż imitujące je maszyny. To androidy poszukujące swego człowieczeństwa, pielęgnują fałszywe wspomnienia, poszukują swej nieistniejącej sfery emocjonalnej, by w końcu popaść w szaleństwo. Ludzie zaś są zdehumanizowani, pozbawieni sfery emocjonalnej, empatii. Tymczasem Lem np. w powieści *Powrót z gwiazd* w masowym zabiegu likwidacji sfery afektywnej u ludzi upatruje szansy likwidacji agresji, konfliktów społecznych itp. Oczywiście powieść bywa różnie interpretowana, wydaje się jednak, że autor *Niezwykłego* z pełną powagą rozpatruje hipotetyczne implikacje zaistnienia takiego panaceum na społeczne bolączki, co musi w dalszych konsekwencjach prowadzić do wizji społeczeństwa cybernetycznego, społeczeństwa–maszyny, społeczeństwa w pełni zrjonalizowanego.

śnienie rozumu, który w końcu staje się ontologiczną podporą rzeczywistości – *logosem*, intelektem „wsiaakającym” w świat, stającym się jego osnową, zasadą istnienia. Ale czy nie jest tak, zadaje niepokojące pytanie Dick, że prawdziwą weryfikacją inteligencji jest jej „zdolność” do szaleństwa? To stany odbiegające od statystycznej normy miałyby być oznaką człowieczeństwa. Być może – mówi nam Dick – jedynie poprzez zanik racjonalności jesteśmy w stanie wyrokować o jej istnieniu (trochę jak czyniła to ongiś teologia negatywna, która poprzez stwierdzenie braku występującej w przyrodzie doskonałości potwierdzała obecność istoty doskonałej). Innymi słowy, to zdolność to szaleństwa może okazać się ostatecznym testem „człowieczeństwa” syntetycznych rozumów – przyznać trzeba, że jest to teza przewrotna. Jak trafnie ujmuje to komentatorka twórczości Dicka, „porównanie androidów do ludzi z zaburzeniami psychicznymi tworzy interesujący punkt widzenia. Podobieństwo między nimi może objawiać się na płaszczyźnie recepcji społeczeństwa. Tak jak szaleńców zamknęliśmy w klatkach i wykluczyliśmy ze społeczeństwa, tak czynimy z androidami. I jedni, i drudzy zagrażają bowiem naszemu porządkowi symbolicznemu, są Innymi” (Radkowska-Walkowicz 2008: 230).

Bibliografia

- Dick P. K. (1998), *Uwagi*, przeł. M. Gawlik, [w:] Tegoż, *Krótki, szczęśliwy żywot brązowego Oxforda*, Prószyński i S-ka, Warszawa.
- Dick P. K. (1999), *Blade Runner. Czy androidy śnią o elektrycznych owcach?*, tłum. S. Kędziński, Prószyński i S-ka, Warszawa.
- Dick P. K. (2000), *Dzień, kiedy pan komputer stracił rozum*, tłum. A. Gawlik, [w:] Tegoż, *Oko Sybilli*, Prószyński i S-ka, Warszawa.
- Dick P. K. (2000), *Elektryczna mrówka*, tłum. M. Gawlik, [w:] Tegoż, *Oko Sybilli*, Prószyński i S-ka, Warszawa.
- Dick P. K. (2000), *Święty spór*, tłum. M. Gawlik, [w:] Tegoż, *Oko Sybilli*, Prószyński i S-ka, Warszawa.
- Dick P. K. (2002), *Nul-0*, tłum. M. Gawlik, [w:] Tegoż, *Czysta gra*, Prószyński i S-ka, Warszawa.
- Dick P. K. (2002), *Ucieczka*, tłum. M. Gawlik, [w:] Tegoż, *Czysta gra*, Prószyński i S-ka, Warszawa.
- Gomułka J., Palm J. (2021), *GOLEM XIV i hierarchia topozoficzna*, [w:] *Filozoficzny Lem. Wybór tekstów Stanisława Lema i opracowania*, F. Kobiela, S. Gomułka (red.), Aletheia, Warszawa.
- Jameson F. (2011), *Archeologie przyszłości. Pragnienie zwane utopią i inne fantazje naukowe*, tłum. M. Płaza, M. Frankiewicz, A. Miszk, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

- Kant I. (1971), *Uzasadnienie metafizyki moralności*, tłum. M. Wartenberg, PWN, Warszawa.
- Kurzweil R. (2013), *Nadchodzi Osobliwość*, tłum. E. Chodkowska, A. Nowosielska, Kurhaus Publishing, Warszawa.
- Lem S. (1964), *Summa technologiae*, Wydawnictwo Literackie, Kraków.
- Lem S. (1971), *Nowa Kosmogonia*, [w:] Tegoż, *Doskonała próżnia*, Czytelnik, Warszawa.
- Lem S. (1976), *Maska*, Wydawnictwo Literackie, Kraków.
- Lem S. (1981), *Golem XIV*, Wydawnictwo Literackie, Kraków.
- Lem S. (1982), *Wizja lokalna*, Wydawnictwo Literackie, Kraków.
- Lem S. (1986), *Rozprawa*, [w:] Tegoż, *Opowieści o pilocie Pirxie*, Wydawnictwo Literackie, Kraków, Wrocław.
- Lem S. (1989), *Fantastyka i futurologia*, t. 1, Wydawnictwo Literackie, Kraków.
- Lem S. (2008), *Niezwyciężony*, Agora, Warszawa.
- Mayer-Schonberger V., Cukier K. (2014), *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, tłum. M. Glatki, MT Biznes, Warszawa.
- Oramus M. (1987), *Biegący po ostrzu*, [w:] Tegoż, *Wyposażenie osobiste*, Iskry, Warszawa.
- Parker J. A. (1992), *(U/od)plciowienie maszyny. „Maska” Stanisława Lema*, tłum. J. Jarzębski, „Teksty Drugie” 3 (15): 93–108.
- Peake A. (2015), *Philip K. Dick. Człowiek, który pamiętał przyszłość*, tłum. T. Hornowski, Rebis, Poznań.
- Penrose R. (2021), *Nowy umysł cesarza*, tłum. P. Amsterdamski, Zysk i S-ka, Poznań.
- Plaza M. (2006), *O poznaniu w twórczości Stanisława Lema*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.
- Popper K. E. (1993), *Spółczeństwo otwarte i jego wrogowie*, t. 1, tłum. H. Krahelska, PWN, Warszawa.
- Radkowska-Walkowicz M. (2008), *Od golema do terminatora. Wizerunki sztucznego człowieka w kulturze*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.
- Rothfork J. (1989), *Duch w maszynie: „Mortal Engines” Stanisława Lema*, [w:] *Lem w oczach krytyki światowej*, red. J. Jarzębski, Wydawnictwo Literackie, Kraków.
- Rozmowy ze Stanisławem Lemem* (rozm. Stanisław Beres) (1987), Wyd. Literackie, Kraków.
- Stoff A. (1983), *Powieści fantastyczno-naukowe Stanisława Lema*, PWN, Warszawa, Poznań, Toruń.
- Szpakowska M. (1996), *Dyskusje ze Stanisławem Lemem*, Wyd. Open, Warszawa.
- Tegmark M. (2019), *Życie 3.0. Człowiek w erze sztucznej inteligencji*, tłum. T. Krzysztoń, Prószyński i S-ka, Warszawa.