

Wprowadzenie

„W maszynie zawsze czaiło się widmo ducha”.

Donna Haraway

18 października 2017 roku świat obiegła wiadomość o niezwykłym sukcesie samouczącego się programu do gry w go opracowanego przez firmę DeepMind. AlphaGo Zero już po trzech dniach samodzielnej nauki był w stanie pokonać swoją wcześniejszą wersję, AlphaGo, wygrywając sto na sto przeprowadzonych gier. Nieco ponad dwa lata wcześniej Google nie kryło satysfakcji, kiedy pierwsza wersja oprogramowania zwyciężyła z Fan Huim, mistrzem Europy w starochińską grę planszową, osiągając przy tym wynik pięć do zera. Kolejne warianty AlphaGo przypieczętowały swoje zwycięstwo w rozgrywce z Lee Sedolem (wygrywając cztery do jednego) oraz z najlepszym graczem na świecie, Ke Jie, nie oddawszy przy tym ani jednej partii. Ulepszony program przyćmił jednak wszystkie osiągnięcia swoich poprzedników.

Trzy dni – zaledwie tyle potrzebował AlphaGo Zero, żeby nadgonić tysiące lat rozwoju wiedzy na temat ludzkiej gry, a oprócz tego wypracować nowe rozwiązania w oparciu o metodę prób i błędów. Jedyne informacje, jakie mu przedstawiono, dotyczyły reguł: pozycji czarnych i białych elementów na planszy. Podczas gdy oryginalny program wyciągał wnioski z milionów wcześniejszych, ludzkich rozgrywek, jego nowy odpowiednik nie otrzymał ani wstępnych uwag, ani żadnych podpowiedzi. System stopniowo ulepszał i udoskonalał swój cyfrowy mózg, wykorzystując do tego wyłącznie własne doświadczenia, to znaczy grając przeciwko sobie. Po trzech dniach treningu Alpha Go Zero pokonał Alpha Go Lee z wynikiem sto do zera,

po czterdziestu zaś zwyciężył z ogromną przewagą (89:11) w starciu z najsilniejszym ze swoich poprzedników, programem Alpha Go Master. Będąc własnym nauczycielem i zarazem przeciwnikiem, AGZ wspiał się z poziomu *tabula rasa* na pozycję najsilniejszego gracza na świecie¹. Było to możliwe prawdopodobnie tylko dlatego, że nie hamowały go żadne ludzkie przedsady ani ograniczenia. Sukces samouczącego się programu został okrzyknięty przełomem w badaniach nad sztuczną inteligencją, a sam AGZ zajął się kwestią zwijania białek, wyświadczając tym samym przysługę współczesnej medycynie.

Czy jeśli maszyna potrzebowała kilku dni na wypracowanie doskonałej strategii, oznacza to, że stopniowo zmierzamy w stronę skonstruowania sztucznej inteligencji, która może przerosnąć nas samych? Temat jest mocno kontrowersyjny, co nie znaczy, że nieistotny – oddech maszyn na karku czujemy przynajmniej od 1997 roku, czyli momentu najsłynniejszego pojedynku szachowego w dziejach. Garri Kimowicz Kasparow, arcymistrz szachowy i wielki przegrany pojedynku z Deep Blue, nie mógł uwierzyć, że przechrzył go bezmyślny komputer autorstwa zespołu IBM. Tę niespodziewaną wygraną odnotowano jako pierwsze w historii zwycięstwo maszyny nad człowiekiem. Nie mniej głośne wydarzenie miało miejsce 7 czerwca 2014 roku, kiedy rosyjski chatbot zaprogramowany przez Władimira Wiesełowa i Eugena Demczenko podołał testowi Turinga i przekonał wymagane trzydzieści trzy procent sędziów, że jest człowiekiem. Szczęśliwy eksperyment przeprowadzono na Uniwersytecie w Reading dokładnie w sześćdziesiątą rocznicę śmierci pomysłodawcy testu. Pomimo pozytywnego wyniku świat zareagował z umiarkowanym entuzjazmem; pojawiły się za to liczne głosy krytyki i rozczarowania. Szczególnie namiętnie wytykano fakt, że Eugene Goostman – zwycięski program – podszywał się pod trzynastolatka z Ukrainy, który wyuczył się angielskiego jako drugiego języka. Co ciekawe, większość internautów pozostała na całe zajście zupełnie obojętna, a niektórzy nie tylko bagatelizowali, ale i podważali sukces Goostmana. Jeden z czołowych transhumanistów,

¹ Por. D. Silver, J. Schrittwieser, K. Simonyan i in., *Mastering the Game of Go without Human Knowledge*, „Nature” 2017, nr 550(7676), s. 354-359.

George Dvorsky, wystosował z tej okazji obszerny komentarz na temat przeceniania testu Turinga; nie omieszczał przy tym przypomnieć, że rosyjskiego chatbota ubiegł trzy lata wcześniej Cleverbot, i że już wtedy nie miało to większego znaczenia².

Idea „gry w imitację” sięga lat pięćdziesiątych ubiegłego wieku i została opracowana przez Alana Turinga – angielskiego matematyka, kryptologa, a także teoretyka sztucznej inteligencji. W klasycznym tekście *Maszyny liczące a inteligencja* zadał on pytanie, czy „maszyny nie mogą wykonywać czegoś, co należałoby nazwać myśleniem, ale co różni się zupełnie od myślenia człowieka”³, i zaproponował rodzaj eksperymentu badającego zdolności komunikacyjne maszyn cyfrowych. Wyznacznikiem sztucznej inteligencji uczynił zdolność do wchodzenia w językową interakcję – sam test miał polegać na pięciominutowej rozmowie, podczas której gracz musiałby rozpoznać, czy rozmawia z człowiekiem, czy jednak z maszyną. Turing – twórca schematu pierwszego komputera – przepowiedział, że za około pięćdziesiąt lat ów gracz nie będzie miał więcej niż siedemdziesiąt procent szansy, żeby dokonać prawidłowej identyfikacji. Przestrzegął przy tym przed zbyt dużym zawieraniem metodzie indukcji naukowej, która ogranicza ludzką wiarę w przyszłe możliwości technologii. To, że do tej pory nie skonstruowano maszyn zdolnych do myślenia i odczuwania emocji, nie przekreśla bowiem jeszcze ewentualności ich powstania.

Wróćmy na moment do naszych komputerowych bohaterów. Dlaczego jesteśmy w stanie uznać zasługi Deep Blue czy Alpha Go Zero, ale nie potrafimy przyznać, że Goostman skutecznie podszycwał się pod człowieka? Program dokonał dokładnie tego, czego chciał Turing – udawał człowieka na tyle dobrze, że doszło do pomyłki. Fakt, że odgrywał rolę nastolatka i oszukał sędziów podstępem,

² Por. G. Dvorsky, *Why The Turing Test Is Bullshit*, <https://io9.gizmodo.com/why-the-turing-test-is-bullshit-1588051412> (dostęp 6 listopada 2017).

³ A.M. Turing, *Maszyny liczące a inteligencja*, przekł. D. Gajkovicz, w: *Maszyny matematyczne i myślenie*, red. E.A. Feigenbaum, J. Feldman, Warszawa 1972, <http://www.kognitywistyka.net/artykuly/at-mlai.pdf> (dostęp 6 listopada 2017), s. 3. Jeśli nie podano inaczej, przekłady na język polski z oryginałów – zarówno w przypadku literatury naukowej, jak i dzieł filmowych – są mojego autorstwa.

przemawia tylko na jego korzyść. Sam tekst *Maszyny liczące a inteligencja* kończy się postulatem zaprogramowania umysłu dziecka, na którym będzie można przeprowadzić odpowiedni proces edukacji. Coś, co jeszcze kilkadziesiąt lat temu przynależało tylko do sfery wyobraźni, dziś traktujemy jako oczywisty i mało imponujący element rzeczywistości – zdajemy sobie sprawę, że za wygraną Goostmana stoją rosyjscy programiści, a skuteczność komunikacyjna wcale nie świadczy o wykształceniu świadomości. Czy więc jesteśmy tak przyzwyczajeni do postępu technologicznego, że eksperyment w Reading nie robi na nas żadnego wrażenia? Czy może boimy się przyznać, że sztuczna inteligencja jest w stanie oszukać własnych konstruktorów, a zatem potrafi coś więcej, niż tylko wygrywać w go czy w szachy? Bezmyślność Deep Blue, Eugene’a Goostmana, a zwłaszcza najnowszej wersji programu AlphaGo wydaje się mocno wątpliwa – maszyny wykonują raczej coś, co (mówiąc po turingowsku) po prostu „różni się zupełnie od myślenia człowieka”. Reakcje na kolejne przełomy wskazywałyby przy tym na to, że cenimy sobie mistrzostwo w planszówkach wyżej od wzajemnego porozumienia, a przynajmniej tylko to pierwsze jesteśmy w stanie zaakceptować. Dopóki komputer zajmuje się czysto rozumową pracą, a metody jego działania dają się zmierzyć i opisać, wszystko jest w porządku. Kiedy jednak nieludzki, dalece niedoskonały program rości sobie prawo do stawiania się na równi z człowiekiem, granica zostaje przekroczona.

Raymond Kurzweil zauważył, że od ponad stu lat moc obliczeniowa maszyn podwaja swoją wartość w ciągu każdych osiemnastu-dwudziestu dwóch miesięcy, i przepowiedział, że za kilkanaście lat będziemy mogli kupić komputery o mocy obliczeniowej zbliżonej do tej ludzkiego mózgu⁴. W pracy *Nadchodzi Osobliwość* zapowiada on przyjście tytułowej Osobliwości Technologicznej, czyli takiego momentu dziejowego, w którym człowiek straci kontrolę nad rozwojem technologii, a ta samodzielnie wyznaczy światu dalszy kierunek ewo-

⁴ Por. R. Kurzweil, *Nadchodzi Osobliwość. Kiedy człowiek przekroczy granice biologii*, przekł. E. Chodkowska, A. Nowosielska, Warszawa 2013, s. 13. W niniejszej monografii używam terminu „Osobliwość” zarówno na określenie hipotetycznego momentu dziejowego, jak i w stosunku do samych przekraczających człowieka sztucznych inteligencji.

lucji. To, co dla Kurzweila i dla innych technoentuzjastów stanowi powód do świętowania, dla innych bywa groźbą ostatecznej degradacji człowieka i nierzadko znajduje odbicie w sztuce jako ucieleśnienie najgorszych ludzkich lęków. Osiągnięcia wymienionych programów świadczą o tym, że sztuczna inteligencja nie jest już jedynie literacką mrzonką; komputerowe programy nie tylko uczą się na własnych błędach, ale też coraz lepiej radzą sobie z komunikacją. Nie myliła się Donna Haraway, kiedy pisała, że „granica między science fiction a rzeczywistością społeczną jest złudzeniem optycznym”⁵.

W niniejszej monografii przyglądam się wybranym tekstom kultury, w których nadejście zapowiadanej przez Raymonda Kurzweila Osobliwości Technologicznej traktowane jest jako poważne zagrożenie dla dotychczasowej pozycji człowieka. Świadoma istnienia przeciwstawnego nurtu, ukazującego maszyny albo z podziwem i pokorą, albo z samym tylko pobłażaniem dla ich służalczego losu, zdecydowałam się na taki klucz doboru tekstów, który wydał mi się płodniejszy interpretacyjnie i pozwolił mi wnikliwiej wykazać złożoność relacji pomiędzy ludzkim a postludzkim. Warto przy tym zaznaczyć, że negatywne (a przynajmniej kontrowersyjne) wizje sztucznej inteligencji pojawiają się w popkulturze o wiele częściej; odpowiednio te pozytywne, a przy tym artystycznie zadowalające, pozostają w zdecydowanej mniejszości (wskazuje to zresztą na niewątpliwe trudności z zaproponowaniem przekonującego wariantu afirmatywnego). Choć zagrożenie ze strony technologii jest już właściwie utartym schematem fabularnym, wciąż trafiają się teksty, którym udaje się w niepozobawiony refleksji sposób uchwycić płynność granicy pomiędzy lękiem a fascynacją wobec własnego wytworu. Źródłem przykładów ilustrujących ową niejednoznaczną obawę przed Osobliwością będą dla mnie filmowe i literackie narracje z gatunku science fiction – te klasyczne, autorstwa Stanisława Lema czy Philipa K. Dicka, te o masowym oddziaływaniu, jak choćby *Matrix* czy *Avengers: Czas Ultrona*, ale i te niemal zupełnie pomijane w dotychczasowych analizach, za to warte uwagi ze względu na sposób

⁵ D. Haraway, *Manifest cyborgów: nauka, technologia i feminizm socjalistyczny lat osiemdziesiątych*, przekł. S. Królak, E. Majewska, „Przegląd Filozoficzno-Literacki” 2003, nr 1, s. 50.

obrazowania sztucznej inteligencji i jej relacji z człowiekiem. Innymi słowy: pochyłę się zarówno nad tekstami, które zajmują znaczące miejsce w popkulturowym obiegu, a przez to organizują wyobraźnię powszechną, jak i tymi, które nie są aż tak popularne, za to w trafny sposób oddają opisywane przeze mnie zjawisko. Pracując nad książką, nie miałam na celu zaproponowania całościowego ujęcia historycznego. Zależało mi raczej na uchwyceniu pewnego fantazmatu, będącego odbiciem czegoś, co można najogólniej określić jako obawę wobec obranego kierunku rozwoju technologii.

Choć wśród badaczy powtarzają się głosy o zmarnowanych szansach literatury science fiction⁶, nawet sceptyczny Andrzej Stoff przyznaje, że jako fakt kulturowy ma ona „konsekwencje świadomościowe: kształtuje, a przynajmniej w dużym stopniu współkształtuje, wyobrażenia o przyszłości”⁷. Jednym z gorliwych obrońców modeli poznawczych generowanych przez science fiction jest Jerzy Jarzębski, którego wypowiedź o Lemie można odnieść także do innych, co ambitniejszych pisarzy czy filmowców:

Nie róbmy z Lema tuzinkowego autora opowieści o latających talerzach. Pomysł jego opakowane są w fantastyczną fabułę, ale w środku tkwi założenie natury filozoficznej, a właściwie – dokładniej – metafizycznej [...]. Apokryficzne wersje fizyki i kosmogonii Lema pokazują nam tedy wizję świata, który rozumnemu nadzorowi i „technologicznemu” kształtowaniu mógłby podlegać – w sensie czysto potencjalnym⁸.

W tej samej monografii pod tytułem *Wszechświat Lema* Jarzębski wysunął tezę, jakoby Lemowy model poznania był „zaklęty w fabułę”⁹. W tekstach kultury, które traktują o hipotetycznych światach i są nakierowane na przyszłość, nie sposób przedstawić złożonej problematyki bez uprzedniego pomyślenia na nowo rzeczywistości – stąd konieczność umieszczenia w książce pobieżnych, często sfragmentaryzowanych streszczeń, które ułatwią czytelnikowi

⁶ Por. Z. Lekiewicz, *Filozofia science-fiction*, Warszawa 1985; A. Stoff, *Czy science fiction może pomóc w rozumieniu rzeczywistości?*, w: „*Metafizyczne*” w literaturze współczesnej, red. A. Koss, Lublin 1992.

⁷ A. Stoff, dz. cyt., s. 62.

⁸ J. Jarzębski, *Wszechświat Lema*, Kraków 2002, s. 98.

⁹ Tamże, s. 118.

orientację wewnątrz danego uniwersum. Z przedstawionych punktów widzenia wyłoni się (post)antropocentryczna wizja świata, który skazany jest na wyniszczające współistnienie ludzkiego i nie-ludzkiego. Światem tym rządzi przemożna obawa – lęk przed kolejnym etapem ewolucji i przed tym, że w zgodzie z odwiecznym prawem natury położy on kres rządóm słabszego ogniwa. Co bardzo istotne, owa fikcyjna rzeczywistość nie jest całkiem wolna od religii czy może raczej od religijnych wyobrażeń. W postludzkiem świecie, przynajmniej dopóki jest on projektowany przez ludzi, Bóg zajmuje swoje stałe, niezagrożone miejsce.

W pierwszym rozdziale omówię podstawowe zagadnienia dotyczące światopoglądu posthumanistycznego oraz przyjrę się negatywnym implikacjom zacieśniania więzi z tym, co nie-ludzkie. Pisząc o metodach przedstawiania technologii jako destrukcyjnej protezy oraz autonomicznego Obcego, powołam się na pierwszy zestaw filmowych i literackich przykładów, a na końcu zasygnalizuję najważniejsze problemy związane z ewentualnym zaistnieniem sztucznej inteligencji. Kolejne rozdziały poświęcę dwóm wybranym wizerunkom Osobliwości Technologicznej obecnym w literaturze, kinie i telewizji: androidom (maszynom antropomorficznym i oblezonym w ciało) oraz superinteligentnym komputerom (zwykle bytom bezcielesnym; jak się okaże, kwestia cielesności odegra tutaj kluczową rolę). Tym samym podążę tropem ewolucji obrazu obcego, technologicznego bytu – od tego, który jest naszą integralną częścią, przez ten manifestujący się jako oddzielne, samowystarczalne ciało, aż po niezależny, przekraczający swój ludzki odpowiednik, niemalże boski umysł. Powtarzalne strukturalne elementy, wykorzystywane przez kolejne teksty poruszające problem nadejścia silnej SI, z czasem dadzą się rozpoznać jako konstytutywne. Dzięki analizie powtarzających się tendencji wykażę związki science fiction z obrazowaniem religijnym i myśleniem numinotycznym oraz spróbuję odpowiedzieć na pytanie, kto i dlaczego przybiera w omawianych tekstach rolę Boga. Nie bez powodu odgrywają ją na przemian konstruktorzy myślących maszyn i same sztuczne inteligencje.

Chciałabym w tym miejscu gorąco podziękować wszystkim, którzy przyczynili się do powstania mojej pierwszej książki, w tym

doktorowi habilitowanemu Andrzejowi Juszczykowi (za niezliczone rozmowy, konsultacje, uwagi i inspiracje), profesor doktor habilitowanej Annie Łebkowskiej (za opiekę naukową i cenną krytykę) oraz doktorowi Sebastianowi Jakubowi Konefałowi (za świeże spojrzenie i wskazanie alternatywnych kierunków myślenia). Osób, którym należą się wyrazy wdzięczności, jest oczywiście znacznie więcej – mogą jedynie mieć nadzieję, że kiedyś odplacę się za całą pomoc i wsparcie, jakiego mi udzielono.

Gwoli wyjaśnienia: w toku argumentacji będę się powoływać zarówno na spostrzeżenia badaczy akademickich, jak i na intuicje twórców kultury (pisarzy i filmowców). Sądzę, że w przypadku opisywanego przeze mnie fantazmatu te dwa dyskursy należy traktować jako komplementarne. Chociaż niniejsza monografia opiera się głównie na analizie fikcyjnych, nierzadko wysoce nieprawdopodobnych narracji, wierzę, że przyczyni się ona do poszerzenia opisu ogólnych mechanizmów kształtujących naszą kulturę, a więc i kierujących naszym myśleniem. Po pierwsze, nie jest tajemnicą, że fantastyka naukowa niejednokrotnie wyprzedziła faktyczny kierunek rozwoju cywilizacji; po drugie, wizje przyszłości pozostają w nieunikniony sposób zdeterminowane przez teraźniejszość¹⁰. Wreszcie – jak podkreśla (za Chadem Oliverem) Grażyna Gajewska – antropologię i science fiction interesują podobne dylematy: „problemy kulturowych kontaktów i kulturowych konfliktów, dyskusje o kulturowym relatywizmie, idea kulturowej rewolucji, cały nacisk w spojrzeniu na rzeczy z różnych perspektyw, pytania o to, co rozumiemy przez społeczeństwo”¹¹. W centrum obu dziedzin tkwi Obcy wraz z metodami jego poznawania i przedstawiania. To, jak ustanowimy Obcego, czyli w tym przypadku maszynę, zadecyduje o naszej własnej pozycji. Tymczasem – jak pisała Haraway – „w prozie, w której żaden bohater nie jest *po prostu* człowiekiem, status człowieka jest bardzo problematyczny”¹².

¹⁰ Por. B. Landon, *Bet On It: Cyber/Video/Punk/Performance*, w: *Storming the Reality Studio*, red. L. McCaffery, Durham, NC 1991, s. 239.

¹¹ G. Gajewska, *Arcy-nie-ludzkie. Przez science fiction do antropologii cyborgów*, Poznań 2010, s. 72.

¹² D. Haraway, dz. cyt., s. 84-85.