



ŁUKASZ ŁAMŻA

POŁĄCZ KROPKI

Nanoboty medyczne,
drony zabójcy, odczytywanie myśli
i inne technologie przyszłości

ŁUKASZ ŁAMŻA

POŁĄCZ KROPKI

Nanoboty medyczne,
drony zabójcy, odczytywanie myśli
i inne technologie przyszłości

Spis treści

Karta redakcyjna

Wstęp

I. Twój robot

I.1. Ciało robota

I.1.1. Robot człękoksztalny

I.1.2. Roje robotów, roboty modułarne

I.2. Umysł robota

I.2.1 Minimum osobowości

I.2.2. Uczucie się

II. Twój świat

II.1. Świat przedmiotów tylko dla ciebie

II.2. Świat treści tylko dla ciebie

II.2.1. Słowa, słowa, słowa...

II.2.2. Obraz, film, muzyka

III. Twoje dane

III.1. Anatomia globalnej maszyny

III.2. Ile jest danych?

III.3. Człowiek i jego dane

IV. Twoje ciało

IV.1. Naprawiamy w nieskończoność

IV.1.1. Inżynieria tkankowa

IV.1.2. Nanocząstki i nanoroboty

IV.2. Projektujemy człowieka idealnego

IV.2.1. Poznaj siebie

IV.2.2. Edytujemy!

IV.3. Człowiek plus

V. Twój umysł

V.1. Mózg-komputer-ciało

V.2. Mózg-komputer (bo po co ciało?)

Na koniec

Spis ramek

Przypisy

© Copyright © by Łukasz Lamża & Copernicus Center Press, 2021

Adiustacja i korekta
Dariusz Niezgoda

Projekt okładki i stron tytułowych
Michał Duława

Skład
MELES-DESIGN

ISBN 978-83-7886-589-6

Wydanie I
Kraków 2021

Copernicus Center Press Sp. z o.o.
pl. Szczepański 8, 31-011 Kraków
tel. (+48) 12 448 14 12, 500 839 467
e-mail: marketing@ccpress.pl
Księgarnia internetowa: <http://ccpress.pl>

Konwersja: [eLitera s.c.](#)

Wstęp

„Nie jest żadną sztuką przepowiadać przyszłość.
Sztuką jest przepowiadać trafnie”.

Andrzej Sapkowski, *Wieża jaskółki*

Masz babo placek: postanowiłem napisać o przyszłości technologii.

Pozwólcie, że króciutko wyjaśnię, jak podszedłem do tego problemu – będzie wtedy łatwiej zrozumieć, jaką właściwie książkę trzymacie w rękach i jak z niej najlepiej korzystać.

Przede wszystkim nie miałem ochoty napisać „wielkiej narracji” na temat przyszłości: książki wizyjnej, odważnej, wieszczącej, co się wydarzy za kilkadziesiąt lat. Prawdę mówiąc, w tego typu tekstach na temat przyszłości, jak choćby *Nadchodzi osobliwość* (Raymond Kurzweil), *Homo deus* (Yuval Harari) albo *Życie 3.0* (Max Tegmark), po prostu brakuje mi konkretów. Miło jest czytać o zgraniu mózgu do komputera, zastąpieniu ludzi przez maszyny albo Superinteligentnym Wielkim Bracie, śledzącym każdy nasz ruch – to atrakcyjne, pobudzające wyobraźnię wizje, w dodatku posiadające jakieś takie oparcie w faktach. Ja jednak, jako dziennikarz naukowy, zawsze zastanawiam się przy okazji tego typu lektur, na ile blisko *tak naprawdę* jesteśmy realizacji tych wizji? Czy jesteśmy od tego o włos, czy też specjaliści, którzy faktycznie pracują przy tych gałęziach technologii, uśmiechają się tylko z politowaniem na fantazje futurologów?

Wizjonerzy lubią bagatelizować „drobiazgi techniczne”, zakładając, że to, co da się wyobrazić, o tym można też pisać – kto zaś zatrzymuje się na detalach, ten nudziarz. Prawda jest jednak taka, że owe „nudne detale” mają decydujące znaczenie dla zrozumienia, co jest możliwe, a co nie. I nie są wcale nudne! Płynna jest granica między futurologią a *science fiction*. Ja chciałem jednak *naprawdę* się zastanowić, co przyniesie przyszłość – nie mogłem więc z czystym sumieniem pisać o wędrownkach w czasie, nadświatelnej podróży kosmicznej i energii próżni. Nawet jednak i bardziej przyziemne marzenia fajnie jest czasem rozłożyć na czynniki pierwsze.

Przykładowo, już od lat 70. XX wieku jesteśmy przekonywani, że „ogólna sztuczna inteligencja” (AGI) – ów mityczny „naprawdę myślący” komputer, coś zupełnie innego od kupki mechanicznych algorytmów – jest tuż za rogiem, za dosłownie dziesięć lat. Jakoś jednak zawsze się okazuje, że dekadę później ów cel znów zostaje odsunięty o dekadę. Dlaczego właściwie tak trudno jest stworzyć AGI, jakiego typu technologie mają największe znaczenie dla rozwoju tego programu, i – skoro już o tym mówimy – co to do kroćset właściwie jest „ogólna sztuczna inteligencja”?! Mój pierwszy cel jest więc taki, żeby trochę te wszystkie wspaniałe technologie odczarować i spróbować wyjaśnić, o co w nich tak naprawdę chodzi.

Z „wielkimi narracjami” jest też inny problem. Zbyt odważne wybieganie w przyszłość może zniechęcać do śledzenia autentycznych postępów w rozwoju tych wszystkich fenomenalnie interesujących technologii. Czytelnicy przyzwyczajeni do „wielkich wizji” mogą być wręcz zawiedzeni po dowiedzeniu się, że szczytem dzisiejszych możliwości jest proteza podpięta do nerwów, ruszająca się zgodnie z intencją ruchową (rozdział V.1), robotyczna ręka ucząca się posługiwania szufelką na podstawie obserwacji zamiatającego człowieka (rozdział I.2.2) albo kawałek wątroby wyhodowany w laboratorium (rozdział IV.1.1). Dlaczego? Ponieważ nie jest to *jeszcze*, odpowiednio, pełne scalenie świata myśli z komputerem, domowy android uczący się jak dziecko albo pełna regeneracja wszystkich tkanek dająca nieśmiertelność. Tymczasem są to wspaniałe, piękne osiągnięcia, kosztujące duże zespoły badawcze lata mrówczego wysiłku, i – jeżeli się nad tym choć przez chwilę zastanowić – balansujące na krawędzi magii! Proteza *podpięta do nerwów*, sterowana *myslą*! Komputer prowadzący debatę, po angielsku, reagujący na argumenty człowieka!

Moim drugim celem jest zarażenie was moim entuzjazmem do wszystkich tych wspaniałych doniesień z ostatnich lat i miesięcy, które stawiają mi włosy na głowie. O ile męczą mnie na dłuższą metę fantaisji, którzy wydają się nie rozumieć omawianych przez siebie technologii, a jednocześnie przewidują ich przyszły rozwój, jeszcze bardziej męczą mnie zblazowane zrzędy. Osobne miejsce w piekle przewiduję dla tych wszystkich malkontentów, na których nie robią wrażenia żadne realne osiągnięcia technologiczne, ponieważ nie spełniają ich abstrakcyjnych wymogów wspaniałości. Choć algorytmy są już dziś w stanie wygenerować fotorealistyczną twarz, która nigdy nie istniała, obraz na dowolny zadany mu temat, jak *Balon nad opuszczonym miastem* albo *Gandalf i Balrog*, chociaż są w stanie

prowadzić debatę na tematy polityczne, tworzyć krótkie opowiadania i artykuły dziennikarskie, muzykę filmową i wygrywać z ludźmi w większość gier, które da się uruchomić na komputerze – zawsze znajdują się chętni, aby przypominać, że nie jest to „prawdziwa” sztuczna inteligencja (której nikt nie umie zadowalająco zdefiniować), a tak w ogóle, to „oprawki okularów na tym obrazie wyglądają trochę nienaturalnie”. To samo zjawisko występuje pośród komentatorów każdego działu nauki i technologii – do dziś podnosi mi się ciśnienie, gdy przypominam sobie narzekania internautów na to, że zdjęcia Plutona wykonane przez sondę New Horizons, są jakieś takie... niskiej rozdzielczości.

To, co mamy do dyspozycji już dzisiaj, teraz, w 2021 roku, jest wspaniałe, niesamowite, piękne, przerażające i fascynujące, a jeżeli ktoś uważa przeciwnie, to się nie zna i niech spada na drzewo.

Stąd metoda, którą przyjąłem.



Być może część z was wie, że od lat czytam na bieżąco artykuły publikowane w najważniejszych czasopismach fachowych świata, między innymi przygotowując swój cotygodniowy przegląd prasy naukowej publikowany na YouTube, zatytułowany *Czytamy naturę*, i redagując dział naukowy „Tygodnika Powszechnego”. Teraz pojawiła się okazja, żeby zebrać to wszystko w jednym miejscu. Zacząłem od jeszcze intensywniejszego niż zwykle wgryzienia się w literaturę naukową, raporty rządowe i pozarządowe, doniesienia z konferencji oraz plotki krążące w światku inżynierów, geeków i dziennikarzy naukowych. Wszystko po to, aby po lekturze tej książki stało się mniej więcej jasne, na jakim etapie rozwoju technologii jesteśmy *tak naprawdę* w 2021 roku.

Ba, postawiłem sobie cel jeszcze ambitniejszy: aby opowiedzieć przy okazji trochę o „zagadnieniach podstawowych”, tak, żeby książka ta mogła służyć jako coś w stylu „podręcznika przyszłych technologii”. Pośród 59 „ramek” rozsianych w tekście tej książki (o których za chwilę) znajdują się także takie, które wyjaśniają podstawy pewnych kluczowych technologii: uczenie maszynowe, rolnictwo komórkowe, edycja genomu metodą CRISPR/Cas, druk 3D, obrazowanie mózgu i wiele innych. Chciałbym, aby książka ta, nawet jeżeli za kilka lat granica rozwoju technologii przesunie się i dzisiejsze cudowne gadzety i wspaniałe wyniki staną się starociami, wciąż mogła się na coś przydać. Moje nadzieje byłyby spełnione, gdyby za 10, a może i 20 lat, ktoś sięgnął po tę książkę i stwierdził, że chociaż jeden z zawartych w niej rozdziałów pomaga mu zrozumieć współczesność.

Teraz parę konkretów.

Książkę tę podzieliłem na pięć rozdziałów wychodzących od pięciu wielkich wizji, przewijających się od lat w popkulturze i mediach:

- I. *Twój robot* – o autonomicznym robocie pomocniku, idealnie cierpliwym, inteligentnym, czułym i domyślnym robotycznym niewolniku na każde nasze skinienie;
- II. *Twój świat* – o świecie sprzętów, mebli, pokarmów, filmów, książek i muzyki powstających tylko dla nas, na każde nasze skinienie, ściśle według naszych oczekiwań;
- III. *Twoje dane* – o całokształcie danych na nasz temat, pracowicie gromadzonych w gigantycznych centrach danych przez wielkie korporacje i bezustannie analizowanych przez sztuczną inteligencję;
- IV. *Twoje ciało* – o zdrowiu i wiecznym życiu za sprawą medycyny jutra, m.in. nanotechnologii, inżynierii tkankowej i edycji genomu;
- V. *Twój umysł* – o scaleniu się naszych mózgów z komputerami.

Każda z tych wizji była dla mnie pretekstem, żeby opowiedzieć wam o podstawach różnych kluczowych technologii („IDEA” ) oraz o stanie rozwoju tych technologii na przykładzie konkretnej realizacji opisanej w ostatnich latach i miesiącach („JEST” ). Stąd rozrzucone w tekście **ramki** – można o nich pomyśleć jako o „cegiełkach przyszłości” albo, w nawiązaniu do tytułu, „kropkach”, które można na różne sposoby łączyć, wyobrażając sobie przyszłość. Książkę najlepiej jest czytać po kolei, nic nie opuszczając – ramki nie są materiałem dodatkowym, tylko kluczową częścią całej opowieści. Mają też jednak pewną samodzielność – a na stronie 277 znajduje się ich pełna lista. Mam nadzieję, że kiedyś w przyszłości – kiedy będziecie czytać o jakimś nowym wspaniałym odkryciu w dziedzinie sztucznej inteligencji, robotyki, nanotechnologii albo inżynierii genetycznej – wpadnie wam do głowy myśl: „Kurczę, była chyba o tym ramka w *Połącz kropki!*”.

Ostatecznie cała przyjemność przewidywania przyszłości spada na was, Czytelników.

Plan jest taki, że ja dostarczam kropki, ale to wy bawicie się w ich łączenie.

Do roboty!

Twój robot

Zacniemy tę książkę od dobrze zdefiniowanego marzenia z długą tradycją: marzenia o robocie pomocniku, maszynie służącej nam na co dzień swoją siłą i cierpliwością, a kto wie, czy nie również domyślnością i intelektem. W idealnych warunkach robot taki powinien być ponadto **androidem**, a więc robotem „człękkształtnym”, przypominającym człowieka. Nazywając rzecz po imieniu, jest to więc marzenie o osobistym służącym, tylko idealnie cierpliwym, wytrzymałym i takim, którego wykorzystywanie nie budzi choćby najdrobniejszych wyrzutów sumienia.

Muszę przyznać, zupełnie szczerze, że nie wygląda na to, abyśmy mieli w najbliższym czasie ujrzyć w naszych domach armie służących nam robotów. Jesteśmy, owszem, otoczeni maszynami i urządzeniami elektronicznymi, które dzielnie nam pomagają w codziennych obowiązkach – a nasze uwikłanie w ów świat pomocnej technologii będzie tylko rość. Już dziś rzeczywistością stają się drony przywożące zakupy oraz autonomiczne odkurzacze i samochody. Tylko od czynników społecznych – bo wiele problemów technologicznych udało już się nam rozwiązać – zależy więc, na ile się upowszechnią. W rozdziale III złożymy wizytę w świecie **internetu rzeczy** – chmury otaczających nas urządzeń domowych, połączonych ze sobą nawzajem i z Internetem, stale analizujących swoje otoczenie, odgadujących nasze życzenia i grzecznie zbierających giga bajty danych na nasz temat.

W tym rozdziale chciałbym jednak trzymać się wizji robotycznego służącego. Nie dlatego, że szczególnie silnie wierzę w jej realizację albo w to, że nawet gdyby do niej doszło, urządzenie takie zmieniłoby znacząco nasze życie. Raczej dlatego, że mam aż trzy niezależne od siebie powody, żeby wejść nieco głębiej w ten temat.

Po pierwsze, ponieważ jest to jedna z tych wizji, które wydają się organizować wysiłek naukowców, twórców i wizjonerów. Przy okazji realizacji tego szaleńczego celu, aby stworzyć

robota na swój obraz i podobieństwo, dokonuje się prawdziwych cudów technologii. Przykładowo, chodzenie na dwóch nogach jest tak naprawdę wybitnie niepraktyczną formą przemieszczania się. Ba, my sami płacimy znaczną cenę za tę ewolucyjną ekstrawagancję – mało kto z czytających te słowa nigdy nie będzie miał problemów z kręgosłupem. Roboty tymczasem mogą mieć dowolną liczbę kończyn, koła, płozy, a nawet śmigła. A jednak – corocznie powstają coraz to lepsze roboty dwunożne, dziś dokonujące już wyczynów prawdziwie akrobatycznych. Inny przykład: twarz ludzka. Nie ma tak naprawdę żadnego dobrego powodu, aby roboty posiadały twarz doskonale naśladującą naszą. Co więcej, budzi to pewien opór estetyczny – o którym trochę jeszcze pomówimy. Starania idące w tym kierunku jednak nie ustają, a na drodze ku realizacji tego celu dokonuje się naprawdę wspaniałych postępów inżynierskich.

Po drugie, ponieważ jesteśmy naprawdę blisko realizacji prostego robota pomocnika. Są obecnie możliwości technologiczne pozwalające na skonstruowanie robota rozmiarów człowieka, który porusza się *względnie* naturalnie, ma *względnie* ludzką twarz, mówi *względnie* ludzkim głosem – i *względnie* z sensem. Gdyby jakiś ekstrawagancki miliarder miał taką fantazję, już dziś mógłby zażyczyć sobie stworzenie robota, który być może nie byłby w stanie jeszcze gotować obiadów albo robić masażu, ale mógłby podać za nim i wykonywać przynajmniej tyle funkcji, ile „inteligentni” asystenci typu Siri, Alexa albo Asystent Google.

Po trzecie wreszcie, ponieważ jest przynajmniej jeden potencjalny rynek dla androidów. Nie zapominajmy, że aby jakaś technologia się upowszechniła – czemu nieuchronnie towarzyszy jej tanienie i doskonalenie – musi być na nią odpowiednio duże zapotrzebowanie. To można oczywiście wytworzyć, jednak znacznie prościej jest opierać się na realnych, organicznych potrzebach ludzkich. Wygląda zaś na to, że robot pomocnik może stać się w przyszłości nie tyle fantazją, co wręcz koniecznością. Myślę o starzeniu się społeczeństw. Wybaczcie mi sporą porcję cyferki zaraz na wstępie tego rozdziału, ale to ważne. (A tych z was, którzy kochają cyferki, informuję, że na początku rozdziału IV znajduje się spora porcja pokrewnych statystyk na temat długości życia ludzkiego).

W 1950 roku osoby w wieku powyżej 70 roku życia – to jedna ze stosowanych dziś, między innymi przez Światową Organizację Zdrowia, definicji „starości” – stanowiły 3% światowej populacji. Dziś jest to 6%, a w 2050 roku^[1*] współczynnik ten ma osiągnąć ponad 11% – 1 osoba na 9 będzie miała 70 lat lub więcej! To jednak średnia światowa, w dużym stopniu kształtowana przez względnie młode społeczeństwa azjatyckie i afrykańskie. W Polsce odsetek osób starych to dziś już 12,3%, a w 2050 roku ma osiągnąć 23% – prawie co czwarta osoba. W Japonii, która od lat jest liderem nie tylko długowieczności, ale i robotyki, osoby stare mają w 2050 roku stanowić 31% populacji tego kraju – czyli będzie to *co trzecia osoba!* Jeżeli macie wątpliwości, czy siedemdziesięciokilkulatki za 30 lat nie będą aby zdrowi i samodzielni, statystyki te można zawęzić do osób w wieku 80+ (16%) albo wręcz 90+ (5%). To wciąż milionowe rzesze Japończyków, z których znaczna część wymagać będzie opieki.

Ludzi w sile wieku będzie tymczasem coraz mniej. Znów garść statystyk, obrazujących, z czym mamy do czynienia. W latach 50. średnia liczba dzieci urodzonych przez jedną kobietę^[2*] przez całe jej życie wynosiła 5. Wieku reprodukcyjnego dożywała z tego mniej więcej czwórka. Dzisiaj w krajach rozwiniętych współczynnik ten wynosi 1,7 (w Polsce 1,5). Wynik poniżej 2 oznacza, nawiasem mówiąc, że społeczeństwo nie „reprodukuje się” – para osób nie „generuje” pary dzieci^[3*]. Mówiąc najprościej: rodzi się coraz mniej coraz dłużej żyjących osób. Trudno przewiduje się przyszłość, ale tego jednego możemy być prawie pewni: XXI wiek będzie epoką staruszków.

Wygląda więc na to, że w przyszłości po prostu zabraknie ludzi młodych, aby zaopiekować się tymi najstarszymi. Zauważmy, że każda pomoc jest na wagę złota – nawet realizacja przez roboty tak „prostych” czynności, jak rozwożenie leków albo wymiana pościeli, stanowiłaby gigantyczną pomoc. Co więcej, badania z 2019 roku na temat naszych preferencji odnośnie do wyglądu „roboty domowego”^[4] pokazują, że to właśnie osoby najstarsze najsilniej domagają się robotów „człekokształtnych”. Młodszy stawiają na ergonomiczny projekt i nowoczesną estetykę, ponad wszystko ceniąc funkcjonalność. Starsi oceniają wysoko ludzką twarz

i naturalny głos, a już najwyżej to, czy z robotem można będzie normalnie porozmawiać, „jak z człowiekiem”.

Czyżby więc androidy pomocnicy mieli stać się w przyszłości nie tyle zabawką miliardarów, co raczej dominującą grupą pracowników w domach spokojnej starości?

Nic to. Bez względu na to, czy wizja ta spełni się, czy nie, póki co robot taki musiałby najpierw powstać. Liczba problemów, które wciąż pozostają do rozwiązania, jest wręcz niebotyczna. Żeby było nam wygodniej o nich myśleć, podzielmy je wstępnie na dwie kategorie: sprzęt i oprogramowanie. Albo – czy wolno już nam tak mówić? – „ciało” i „dusza”.

I.1. Ciało robota

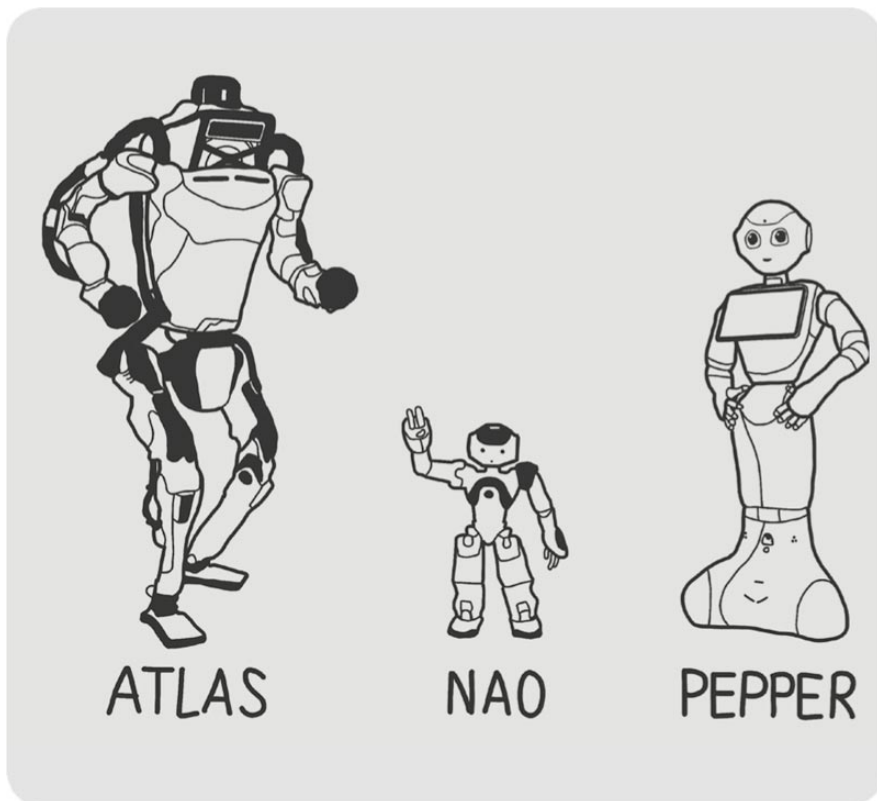
Fundamentalne pytanie, które musimy sobie zadać, gdy planujemy robota domowego, jest takie: czy musi on koniecznie przypominać człowieka – a jeśli tak, to w jakim stopniu? Każde rozwiązanie, jak to w inżynierii, ma swoje wady i zalety.

I.1.1. Robot człekokształtny

Wszystkie badania pokazują, że ludzie reagują znacznie lepiej na maszyny człekokształtne. Jeśli zaś mówimy o „moim robocie” – pomocniku, towarzyszu, przyjacielu – to z góry zakładamy, że będzie to raczej *ktoś* niż *coś*. Istnieją już dzisiaj produkowane masowo „standardowe” roboty humanoidalne, wykorzystywane powszechnie przez naukowców w sytuacjach eksperymentalnych. Jednym ze światowych liderów w produkcji robotów humanoidalnych jest SoftBank Robotics (wcześniej pod nazwą Aldebaran Robotics). To z ich „stajni” wyszły między innymi niewielki dwunożny Nao (60 cm wzrostu, 5,5 kg) i nieco większy, „półhumanoidalny” Pepper na kółkach (120 cm, 28 kg). Roboty te cechują się prostą, miłą dla oka anatomią, są wyposażone w zestawy gotowych czujników przyzwoitej jakości i „współpracują” z popularnymi językami programowania, dzięki czemu ich „mózgi” względnie łatwo rozbudowywać.

Szybko rozwija się też rynek robotów „supersprawnych”. Boston Dynamics, firma wykonująca m.in. roboty o potencjale militarnym dla amerykańskiej agencji badawczej DARPA, przedstawiła w 2013

roku dwunożnego ATLAS-a (150 cm, 80 kg), który w 2019 roku potrafił już stać na rękach i wykonywać salta. W grudniu 2020 roku dwa Atlasy pochwały się całemu światu swoimi kocimi ruchami, tańcząc – nieźle, a w każdym razie na pewno lepiej niż ja – w rytm *Do you love me*. Inna wiodąca na rynku firma, Agility Robotics, pracuje nad dwoma robotami – Cassie i nowszym Digit – które ruszają się sprawnie po miejskim terenie, wchodząc na przykład bez trudu po schodach i radząc sobie odruchowo z nierównymi chodnikami. Wiosną 2021 roku MIT ogłosił^[2] pierwsze wyniki badań nad swoim „ultrasprawnym” robotem, stworzonym z myślą o akrobatyce i parkour.



Człękkształtna budowa ciała ma oczywiście wiele wad z inżynierskiego punktu widzenia. Poruszanie się na dwóch nogach jest kłopotliwe, trudne i niepewne: nawet dorosłemu, wysportowanemu człowiekowi zdarza się czasem potknąć i stracić równowagę we własnym domu. A co, gdyby przydarzyło się to 80-kilogramowemu Atlasowi przynoszącemu nam właśnie kawę? Dużo

pewniejsze i łatwiejsze w zaprogramowaniu są cztery nogi albo choćby, jak u Peppera, „prawie nogi”, czyli struktura z grubsza przypominająca ludzkie kończyny dolne, ale zakończona kółkami.

Po co więc w ogóle wikłać się w dwunożność? Wygląda na to, że budowa taka jest zaskakująco uniwersalna, zwłaszcza w przestrzeni naszych domów i miast, które powstały wszak dla istot dwunożnych. Istnieją oczywiście roboty czworonożne, które umieją już całkiem nieźle wdrapywać się na schody – na przykład Spot od Boston Dynamics – te jednak z kolei nie powtórzą innego słynnego wyczynu Atlasa, jakim było wejście do samochodu i kierowanie nim, po prostu z fotela kierowcy. Oczywiście, funkcję tę można by zrealizować na wiele innych sposobów: bądź co bądź, skoro mówimy już o przyszłości technologii, trudno zapomnieć o samochodach autonomicznych, które w ogóle nie potrzebują kierowcy, ludzkiego czy robotycznego. Nie da się jednak zaprzeczyć, że dwunożny, antropoidalny robot to rozwiązanie o tyle „eleganckie”, że jego właściciel nie musiałby się co chwilę zastanawiać, czy powierzone mu zadanie jest dla niego wykonalne: wystarczyłoby wyobrazić sobie, że wydajemy dyspozycje człowiekowi, posługując się przy tym zwykłą intuicją. Jeśli interesuje nas całkowita „przezroczystość” tej technologii – czyli robot rzeczywiście płynnie wtapiający się w tok naszego życia – postać androida to najbardziej naturalne rozwiązanie.

Zacznijmy może od pytania, jak bardzo podobny do człowieka musiałby być właściwie robot, abyśmy potraktowali go z sympatią i zgodzili na jego obecność w życiu codziennym. Okazuje się, że potrafi nas zadowolić nawet najprostsza forma „uczłowiczenia”.



Poćwicz ze mną!

Dużą częścią robotyki są dzisiaj badania społeczne i psychologiczne relacji człowiek-robot. Co właściwie decyduje o tym, czy potraktujemy daną maszynę jako partnera? Co sprawia, że utożsamiamy się z robotem, przypisujemy mu emocje, czujemy przyjemność z interakcji z nim?

Badania pokazują^[3], że nawet tak elementarna sztuczka, jak patrzenie rozmówcy w oczy przez humanoidalnego robota – co czasem sprowadza się do

wyświetlania plamki reprezentującej oko w nieco innym miejscu – znacząco zwiększa sympatię, jaką ludzie odczuwają względem niego. Ba, wykonywano eksperymenty^[4] z robotami oprowadzającymi po muzeum, w których celowo zaprogramowano „faworyzowanie” jednego z gości poprzez patrzenie mu w oczy częściej niż w oczy innych. Okazuje się, że osoba „wyodróżniona” przez robota lubiła go bardziej, a pozostali uczestnicy wycieczki wykazywali ślady „zazdrości”.

W 2017 roku opisano^[5] prosty eksperyment, w którym robot Nao posłużył jako „fitness coach” w domu opieki dla osób starszych. Najpierw sam Nao został nauczony szeregu ćwiczeń przez trenera fitness, obserwując go i próbując powtórzyć jego ruchy. To ważny etap (już sam w sobie dość wymagający inżynieryjnie), dzięki któremu późniejsze nauczanie robota nowych ćwiczeń sprowadzałoby się do kolejnej sesji z trenerem i nie wymagało „grzebania w kodzie”.

Robota wpuszczono następnie do domu opieki, poświęcając osobne pomieszczenie na „salę ćwiczeń z robotem”. Uczestnikami było 12 osób w wieku 70–88 lat, którym pozwolono wybrać, czy mają ochotę na zwykłą sesję treningową, czy na eksperymentalne spotkanie z robotem. Chęć taką wyrazili wszyscy zapytani – ewentualne lęki i wątpliwości pokonała najwyraźniej przemożna chęć zerwania z rutyną. Większość uczestników wzięła też udział w pełnej serii trzech spotkań. Robot całkiem nieźle radził sobie w trybie autonomicznym, podczas którego pozostawiony był sam na sam z uczestnikami – liczba przypadków, kiedy uczestnicy wzywali ludzkich opiekunów w trakcie sesji, szybko spadała. Przy trzecim spotkaniu potrzebna już była średnio tylko jedna interwencja człowieka na 10-minutową sesję treningową.

Zacznijmy od wyników praktycznych. Robot miał dostęp do kamery z „sonarem”, a więc rejestrującej równocześnie obraz i odległość do przedmiotów: dzięki temu mógł rekonstruować pozę przyjmowaną przez osobę ćwiczącą i reagować na nią. Jeżeli ćwiczenie było wykonywane prawidłowo, Nao udzielał pochwały. Nieprawidłowości były korygowane w odniesieniu do „modelu idealnego” stworzonego w oparciu o zarejestrowane wcześniej ruchy instruktora. Cytowany tu artykuł szczegółowo omawia skutki ćwiczeń – zasadniczo pozytywne – jednak bardziej interesujące są dla nas w tym momencie kwestie relacji człowiek–robot.

Nao wyposażono tylko w trzy proste funkcje „ocieplające wizerunek”:

- automatyczne wodzenie głową za człowiekiem,
- mruganie oczami (poprzez włączanie i wyłączenie diody reprezentującej oczy),
- używanie imienia osoby ćwiczącej.

Robot ten jest fabrycznie wyposażony w moduł rozpoznawania emocji, przyporządkowujący do obrazu z kamery jedną z sześciu etykiet: zadowolony, zaskoczony, smutny, znudzony, zirytowany i zagubiony (czyli z miną wyrażającą niezrozumienie). Osoby tworzące „swojego” robota na bazie Nao muszą dopiero zaprogramować reakcję na te emocje. Algorytm^[4*] opisany w artykule przewidywał bardzo proste sposoby reagowania, na przykład powtórzenie poprzedniego komunikatu w reakcji na zagubienie albo komunikat pozytywny w reakcji na ziryutowanie. Aha, robotowi nadano też imię Leyla i ubrano go w różowy sweterek.

Wyniki przeprowadzonej po eksperymencie ankiety są szalenie interesujące. Przede wszystkim niemal wszyscy uczestnicy zgłosili bardzo wysoki poziom zadowolenia z sesji i z kontaktu z robotem (wyjątkiem było dwoje uczestników,

którzy ze względu na poważne problemy ze słuchem nie rozumieli dobrze poleceń robota i wskutek tego otrzymywali wiele komunikatów „negatywnych” – nawet oni jednak oceniali samą sesję jako przyjemną, a tylko robota – nisko). Wykazywali też spontanicznie próby nawiązania głębszego kontaktu i inicjowali rozmowę – niestety, w tym eksperymencie Nao nie został wyposażony nawet w prosty moduł konwersacyjny.

Na pytanie, czy zgodziliby się na to, aby ich trener był robotem, uczestnicy odpowiadali wręcz entuzjastycznie (średnia powyżej 9 w skali od 1 do 10), natomiast ze znacznie większą rezerwą reagowali na pytanie, czy robot mógłby być ich przyjacielem (zakres wyników 1,5-10 przy średniej nieco poniżej 7 – czyli i tak względnie wysoko!). Oto niektóre komentarze uczestników:

- „Ludzie mi wystarczą za przyjaciół”.
- „Wolę prawdziwą, bliską, realną przyjaciółkę, jak moja sąsiadka”.
- „Robot nie może zastąpić człowieka, a posiadanie robota przyjaciela mogłoby osłabić moje relacje z ludźmi”.
- „Posiadanie robota przyjaciela mogłoby mnie odizolować od społeczeństwa”.
- „Robot nie potrafi zrozumieć znaczenia moich słów, więc jak mogę do niego mówić? A bez rozmowy nie może być moim przyjacielem”.

Dłuższe wywiady ujawniły też jednak dodatkowe motywacje. Uczestnicy eksperymentu, zapytani o to, czy chcieliby posiadać tego typu robota w domu jako pomocnika, zwykle zaprzeczali. Czemu jednak? „Nie chcę brać na siebie odpowiedzialności za robota”. „Ładowanie i naprawianie robota to dodatkowe obciążenie”. „Boję się, że stanę się od niego zależna”.

Jakie płyną wnioski z tego eksperymentu? Po pierwsze, jak zawsze, szalenie istotne okazują się „przyjemne” względy praktyczne: dobrze dobrany poziom głośności, łatwość ładowania. Dopóki urządzenie nie będzie dobrze, ergonomicznie zaprojektowane, nie pomoże nawet najbardziej wyrafinowana sztuczna inteligencja. Po drugie, ta może nie być w ogóle potrzebna. To niezwykle, jak przy pomocy kilku prostych sztuczek – jak mruganie oczami albo wodzenie wzrokiem – robot „nabiera życia” w oczach człowieka. Na myśl przychodzi słowo „pareidolia”: skłonność do doszukiwania się twarzy ludzkiej w chmurach czy korze drzewa. Nasz mózgowy „obwód poszukiwania człowieka” jest naprawdę niezwykle czuły i nawet na przekór temu, co mówią nam oczy i rozum, bardzo *chcemy* nadać maszynie osobowość i człowieczeństwo. Po trzecie, dobrą ścieżką wejścia robotów w świat człowieka jest prosta, wąsko określona rola. Dużo łatwiej jest wyobrazić sobie robota trenera fitness, który niejako przy okazji swojej podstawowej funkcji zaczyna z czasem wchodzić ze swoim klientem w coraz głębsze rozmowy, niż robota od początku reklamowanego jako „przyjaciela”. To zaś, swoją drogą, nie jest takie zaskakujące, biorąc pod uwagę bardzo podobne mechanizmy w relacjach człowiek-człowiek. Wiele pięknych przyjaźni rozpoczęło się od przelotnej relacji zawodowej albo usługowej.

Oczywiście, w pewnych zastosowaniach doskonała iluzja człowieczeństwa może się okazać konieczna. Jednym z takich zastosowań są **sekslalki**, których twórcy nalegają, by ich produkty określać jako seksroboty. Rynek ten rozwija się z niebywałą dynamiką,

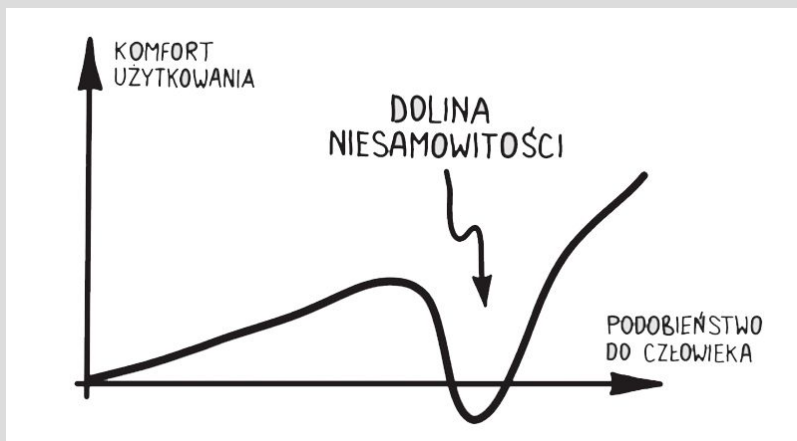
wychodząc naprzeciw oczekiwaniom klientów, dla których podobieństwo do żywego człowieka jest na wagę złota. Firma RealDoll oferuje obecnie między innymi model Harmony^[54], pokryty miękkim polimerem udającym skórę, który można za opłatą wyposażyć w piegi i inne drobne niedoskonałości cery. Robot wodzi wzrokiem za ludźmi, mruga i zmienia wyraz twarzy dzięki ukrytym pod powierzchnią skóry siłownikom. Zaprzyjaźniona, choć też i konkurencyjna, firma Realbotix oferuje wymienne głowy dla Harmony, wyposażone również w moduł konwersacyjny, dzięki czemu z seksrobotem można prowadzić prostą rozmowę. W listopadzie 2020 roku przedstawiciele tej firmy ogłosili^[6], że pracują nad w pełni „samodzielnym”, tj. autonomicznie poruszającym się seksrobotem, którego cenę szacują na 30–60 tysięcy dolarów.

Prowadzi się też prace nad androidami do obsługi klientów, na przykład w hotelach albo na targach przemysłowych. Od 2003 roku nad długą serią takich robotów pracuje japońska firma Kokoro, a stosowana przez nich nazwa **Actroid** bywa używana jako ogólne określenie na androidy o tak realistycznych twarzach, że „mogłyby być aktorami”. Południowokoreański instytut badawczy KITECH przedstawił też cztery coraz to bardziej złożone androidki o imieniu EveR, z których najbardziej zaawansowana EveR-4 miała już twarz poruszaną od wewnątrz przez 30 punktów przyczepu. Przyglądając się dostępnym materiałom, sądzę, że odpowiednio roztargniony gość hotelowy mógłby w środku nocy wejść do recepcji, załatwić nocleg i udać się do pokoju, nie zorientowawszy się, że miał do czynienia z tego typu androidem.



Dolina niesamowitości

Trudno nie wspomnieć przy tej okazji o **dolinie niesamowitości** (*uncanny valley*) [6*]. Wyobraźmy sobie wykres przedstawiający na osi poziomej stopień podobieństwa danego robota do człowieka, a na osi pionowej stopień komfortu lub przyjemności z obcowania z nim:



Teoretycy, a następnie i praktycy w obszarze robotyki zauważyli, że w miarę upodabniania się robota do człowieka rośnie komfort przebywania w jego obecności, jednak na poziomie tuż przed osiągnięciem doskonałego podobieństwa występuje potężna „dolina” w wykresie. Obcowanie z robotem, który *prawie* wygląda jak człowiek, budzi nieprzyjemne odczucia – podobnie zresztą jak obcowanie z człowiekiem o zniekształconej twarzy, który w związku z tym również nie do końca „przypomina człowieka”. Występowanie tej dolinki potwierdzono niedawno w badaniach empirycznych [7] – ludzie rzeczywiście wydają się mniej lubić robota wyglądającego „prawie jak człowiek” niż takiego, który ma w miejscu twarzy po prostu plastikowy cylinder z oczami. Dopiero *naprawdę dobra* iluzja człowieka budzi pozytywne emocje: ponieważ nasz mózg zaczyna wtedy traktować androida po prostu jak człowieka.

Prowadzi to do paradoksalnej sytuacji, w której twórcy robotów muszą czasem celowo podkreślać, że ich twory to maszyny, na przykład – to popularna sztuczka – wykonując czaszkę z przezroczystego plastiku, pod którym znajdują się rozmaite, niekoniecznie pełniące realne funkcje, elementy elektroniczne i „mrugające światełka”.

Twarz, skóra i mimika to jedno. Dużą przeszkodą na drodze ku przeskoczeniu doliny niesamowitości jest płynność ruchów. EveR-4 wygląda naprawdę nieźle, jednak sposób, w jaki przekręca szyję, nie ma w sobie nic ludzkiego: widać gołym okiem, że siedzi tam w środku po prostu silniczki wykonujące zaprogramowaną liczbę obrotów.

Wszyscy pamiętamy ze starych filmów science-fiction nienaturalne, mechaniczne ruchy robotów takich jak C-3PO, Robocop czy nawet Terminator. Choć kinematografia nadąża za rozwojem naszej fantazji

i z każdą kolejną dekadą filmowe roboty ruszają się coraz bardziej jak człowiek, rzeczywista robotyka postępuje w jej ślady wolniej. Tymczasem naturalne, dające poczucie bezpieczeństwa, „ludzkie” ruchy są równie istotne, co realistyczna anatomia i ludzka twarz.

Nie jest łatwo nauczyć robota, aby wykonywał płynne, organiczne ruchy, zwłaszcza w nowym środowisku. Roboty przemysłowe, lutujące układy elektroniczne, montujące silniki albo rozwożące towary po magazynie, pracują w cieplarnianych warunkach. Przedmioty, z którymi oddziałują, są im podtykane pod nos z milimetrową precyzją, a wszystkie półki w magazynie oznaczone są własnymi kodami QR. Po wpuszczeniu ich do przydomowego warsztatu albo do lasu stałyby się jednak bezradne jak niemowlęta. Jeżeli jednak chcemy, żeby „mój robot” radził sobie w realnym świecie, w którym na podłodze może leżeć kocyk, a rączka patelni jest utracona albo tłusta od oleju, musi być zdolny do samodzielnego, elastycznego planowania i wykonywania ruchu. Ładniej i precyzyjniej: „kinetycznej improwizacji”. Przyjrzyjmy się postępowi w tej dziedzinie, na warsztat biorąc elementarną, ale jakże istotną umiejętność chodzenia.

„Klasyczne” podejście do problemu chodzenia przedstawia się następująco. Najpierw wymyślamy, jaką metodą ma się przemieszczać nasz robot. Przypuśćmy, że mamy do dyspozycji cztery nogi i po namyśle stawiamy na chód „à la stęp konia”. Programujemy więc odpowiednią sekwencję ruchów, zwykle silnie inspirując się przykładami biologicznymi (lewe tylne udo o 15 stopni w przód, skręt w biodrach, prawe kolano o 5 stopni wstecz...) i... testujemy. Nasz robot oczywiście pada jak długi albo, w najlepszym razie, buja się w miejscu. Dlaczego? Cóż, nawet nieznaczna nierówność terenu, wiatr albo niesymetryczne obciążenie grzbietu sprawiają, że wymyślone przez nas „na sucho” 15 stopni skrętu tylnego uda okazuje się fikcją. Faszerujemy więc naszego robota czujnikami, które rejestrują, kiedy dochodzi do kontaktu każdego „kopytka” z podłożem, jaki jest ogólny stan równowagi całego ciała, siła nacisku i kąt w poszczególnych stawach – i tak dalej. Mało tego, możemy też przecież wyposażyć naszą mechaniczną bestię w zmysł wzroku.

I tutaj zaczyna się magia – musimy bowiem przetłumaczyć te informacje na komendy „korygujące” ruch. Najładniej chodzące i najsprytniej reagujące na wyzwania roboty, jak choćby BigDog i Spot wspomnianej wyżej firmy Boston Dynamics – które są już w stanie biegać po nierównym, śliskim albo kamienistym terenie i utrzymać równowagę po solidnym kopniaku w bok – są produktem wielu lat żmudnego programowania przez sztaby inżynierów,

współpracujących nierzadko z zoologami, biomechanikami i fizjologami sportu. Boston Dynamics nie ujawnia algorytmów swoich robotów, jednak na podstawie rozmaitych wystąpień konferencyjnych i artykułów naukowych autorstwa pracowników tej firmy^[8] można zrekonstruować ich zasadniczy *modus operandi*. Idea jest taka, aby stworzyć jak najlepszą „teorię chodu”, a następnie przetłumaczyć ją na mechanizmy kontroli ruchu; co robić przy zmieniającym się nachyleniu terenu, co robić, gdy nogi zaczynają się ślizgać, co robić, gdy grunt jest grząski... tysiące, tysiące linii kodu i parametrów.

Projektowanie chodzących robotów – oraz, analogicznie, robotów przynoszących kawę albo zmywających naczynia – poprzez „ręczne” nasycanie ich wiedzą ekspercką to skuteczna, choć kłopotliwa, metoda i nieprędko odejście do lamusa. Ktoś naiwny mógłby zapytać: a czy nie dałoby się prościej?

Ów ktoś chyba miałby rację.



Uczenie przez wzmacnianie

To chyba dobra okazja, żeby wprowadzić kilka elementarnych pojęć związanych z zagadnieniem „**sztucznej inteligencji**” (*artificial intelligence*, AI). O samej „sztucznej inteligencji” będę w tej książce mówił niewiele, ponieważ pojęcie to jest tak naprawdę gigantycznym workiem o mętnych granicach, w którym mieszczą się dziesiątki najróżniejszych metod i sztuczek. Ich jedyną cechą wspólną jest to, że mają one odtwarzać (a zwykle: udawać) przejawy naszej własnej, ludzkiej inteligencji: od rzeczy tak elementarnych, jak sprawne chodzenie, przez rozwiązywanie krzyżówek, rozumienie mowy, aż po tworzenie muzyki. Teoretycy AI lubują się w wyklócaniu, czy dany algorytm to „już” sztuczna inteligencja czy „jeszcze” tylko kilka linijek kodu – trochę więcej wyzłóśliwiam się na ten temat na samym początku rozdziału 1.2. My w tej książce nie będziemy bawić się w teoretyzowanie i zamiast o „sztucznej inteligencji”, będę raczej mówił o konkretnych technologiach, które przepełniają dziś świat nauki i inżynierii. Niniejsza ramka to pierwsze z wielu miejsc, gdzie przedstawię niektóre spośród tych technologii.

Aby zrozumieć ideę **uczenia przez wzmacnianie** (*reinforcement learning*), wystarczy zastanowić się, w jaki sposób my sami nauczyliśmy się chodzić albo mówić, albo robić jajecznicę. Nad wykonującymi pierwsze kroki bobasami nie stoi sztab fachowców, podpowiadających, z jaką siłą należy spinać mięśnie uda, gdy noga natrafia na miękki materacyk, a z jaką, gdy jest to twarda podłoga. Uczymy się spontanicznie, reagując na to, co mówią nam nauczyciele i co „mówi” nam świat.

Komputery były początkowo „maszynami programowalnymi”, których algorytmy były pieczołowicie projektowane, komenda po komendzie. Jeden z wielkich geniuszów i wizjonerów XX-wiecznej informatyki, Alan Turing, już w 1950 roku zaproponował^[2] model komputera, który samodzielnie generuje algorytmy, ucząc się na swoich błędach i sukcesach. W II połowie XX wieku program ten był stopniowo rozwijany, ale dopiero w XXI wieku udało się przemienić go w skuteczne narzędzie, będące jednym z fundamentów programowania. Uczenie przez wzmacnianie jest, nawiasem mówiąc, jedną z gałęzi **uczenia maszynowego** (*machine learning*). Termin ów, który będzie regularnym gościem na kartach tej książki, obejmuje wszystkie przypadki, gdy algorytmy komputerowe automatycznie stają się coraz skuteczniejsze wskutek nabierania „doświadczenia” i analizy danych.

Jak mogłoby wyglądać uczenie chodzenia przez wzmacnianie? Okazuje się, że gdybyśmy chcieli wcielić ten plan w życie w najprostszy, „naiwny” sposób, szybko pojawiłyby się problemy praktyczne o kolosalnej trudności. Plan taki mógłby wyglądać następująco: Zaczynamy od zupełnie „naiwnego” robota, karmionego wyłącznie surowymi danymi z rozsznianych po jego ciele czujników. Następnie każemy mu poruszać się w zupełnie przypadkowy sposób – dane z czujników są tłumaczone na polecenia dla kończyn na rozmaite, eksperymentalne sposoby. Z reguły owym algorytmem tłumaczącym jest **sieć neuronowa** – struktura obliczeniowa inspirowana budową zwierzęcego mózgu. O sieci neuronowej można pomyśleć w pierwszym przybliżeniu jako o kłębowisku komórek (neuronów), które w odpowiedzi na „bodźce” (dane wpływające, czyli „wejście”) generują jakies „zachowanie” (dane wypływające, czyli „wyjście”). Jeśli któryś neuron będzie nieco inaczej reagował na wpływające do niego sygnały, to zmieni się sposób reakcji całej sieci. Odpowiednia złożona sieć neuronowa bywa czasem porównywana do „czarnej skrzynki”, która w reakcji na „wejście” (może być to zbiór liczb, ale też na przykład dźwięk albo obraz) produkuje „wyjście” (na przykład liczbę albo słowo, albo zbiór poleceń dla „mięśni”).

Przypuśćmy, że zaczynamy od naprawdę „naiwnej” sieci neuronowej, która dane z czujników przetwarzać będzie na sygnały dla „mięśni” w zupełnie przypadkowy sposób. Jeżeli wyobrazimy sobie, że każdy neuron to kalkulator, który w reakcji na liczbę wykonuje ustaloną sekwencję operacji, to neuron „naiwny” powstanie, gdy sekwencję tę ustalimy poprzez rzuty kością do gry.

Niemal zawsze skutkiem działania tego typu sieci będzie szpatmatyczne drganie albo bezruch – czasem jednak, czystym przypadkiem, robotowi może się udać podpełznąć kawałek do przodu albo wzniesić na kolana. Takie przypadki „nagradzamy” możliwością „rozmażania się”: następne pokolenie robotów powstaje na podstawie „genów” – czyli opisu tego, w jaki sposób tłumaczone są sygnały z czujników na ruch – najlepiej chodzących osobników, poddanych lekkiej modyfikacji. Taki model uczenia byłby też więc jednocześnie **algorytmem genetycznym**. Okazuje się, że tego typu metoda, opierająca się na szukaniu po omacku i wzmacnianiu nawet najdrobniejszych sukcesów, jest zaskakująco skuteczna. Istnieją w internecie proste symulatory, ilustrujące nabywanie umiejętności chodzenia przez początkowo przypadkowo reagujące roboty^[7*].

Choć sama idea jest słuszna i jest dziś podstawą dla milionowych zastosowań, mamy pewien problem z planem, aby tą metodą nauczyć robota chodzić: naprawdę „naiwne” algorytmy zwykle potrzebują sporo czasu, żeby wykonać – dosłownie lub w przenośni – pierwszy krok. Mówimy tu o setkach, tysiącach albo wręcz dziesiątkach tysięcy „pokoleń”, wykonujących kompletnie bezsensowne, czasem wręcz samobójcze ruchy, zanim przypadkiem uda się natrafić na „obiecujący trop”, później względnie szybko wzmocniony do postaci dobrego rozwiązania.

Gdy „świat”, którego chcemy nauczyć maszynę, sprowadza się do szachownicy, możemy bez większych kosztów trenować tysiące „uczniów”, co sekundę każąc im rozgrywać kolejną partię szachów. Co więcej, gdy po kilku tysiącach pokoleń okaże się, że ewolucja utknęła w martwym punkcie, można szybko dokonać drobnej modyfikacji algorytmu uczącego i rozpocząć od nowa. Gdy jednak mowa o rzeczywistych kroczących robotach, koszt takiego przedsięwzięcia i czas jego trwania szybko stają się absurdalne. Tysiąc „głupio” chodzących robotów, miesiącami tłukących bez ładu kończynami o podłozę, to koszmar inżyniera – i księgowego. Mało tego, jest też inny problem: możemy sobie zażyczyć, aby nasz algorytm optymalizował nie tylko „software”, ale i „hardware” robota. Tak przecież działa prawdziwa ewolucja: pierwsze kręgowce spacerujące po lądzie nie tylko uczyły się nowego sposobu machania płetwami, ale też przekształcały je stopniowo w „prawdziwe” kończyny. Gdyby więc nasz algorytm „podpowiadał” nam modyfikacje długości kończyn albo zakresu ruchliwości stawów, koszmar inżyniera – i księgowego... – wkroczyłby na zupełnie nowy poziom makabry.

Można przyjąć dwie zasadnicze strategie. Pierwsza to rozpocząć od stanu „nie całkiem naiwnego” – tak, aby już pierwsze pokolenie uczących się chodzić robotów chodziło przynajmniej „nieźle”. Oznacza to w praktyce, że rola uczenia maszynowego sprowadzona zostaje do „dopieszczenia” algorytmów stworzonych metodą „tradycyjną” (ekspercką). To oczywiście i tak się już dzieje. Nawet „tradycyjni inżynierowie” bardzo chętnie posługują się dziś uczeniem maszynowym przy optymalizowaniu ruchu swoich maszyn kroczących, których zasadnicza „logika” ruchu opracowana została metodą ekspercką.

Przypuśćmy jednak, że chcemy „iść na całość” i stworzyć maszynę, która nauczyłaby się chodzić, startując zupełnie „od zera”. Tu pojawia się druga metoda: „trening wirtualny”. Gdyby udało się stworzyć odpowiednio realistyczny model robota oraz odpowiednio

realistyczny model świata, to *czysto teoretycznie* można by przecież rozpocząć od treningu w wirtualnej rzeczywistości, a dopiero potem przekazać uzyskaną w ten sposób wiedzę – i anatomie! – prawdziwym, fizycznym robotom. Plan taki wyłonił się mniej więcej w połowie drugiej dekady XXI wieku, a w okolicach 2018 roku pojawiły się jego pierwsze udane realizacje, głównie w warunkach laboratoryjnych. A oto jeden z pierwszych przypadków, opisany w październiku 2020 roku, kiedy metodą taką udało się wytrenować naprawdę zręczną bestię.



Sam się nauczę chodzić^[10]

Opisywany w tej pracy robot powstał w Robotic Systems Lab w Zurychu. Trenowany był przytoczoną wyżej metodą „najpierw symulacja – potem rzeczywistość”. Najpierw więc stworzono wyrafinowaną symulację komputerową świata fizycznego, a następnie umieszczono w nim równie realistyczny model robota i kazano mu skutecznie przemieszczać się po różnorodnym terenie. Z każdą kolejną próbą „mózg” robota (formalnie: po prostu „algorytm”) stawał się coraz lepszym kontrolerem jego kończyn. Dopiero po osiągnięciu zadowalającej skuteczności ten sam „mózg” skopiowano do rzeczywistego, fizycznego robota.

Najpierw przetestowano go w warunkach laboratoryjnych na względnie prostym terenie znanym już z symulacji komputerowych, a następnie również na zupełnie nowych typach podłoża, na przykład śniegu albo omszałych pniakach. Okazało się, że mechaniczny czworonóg doskonale radzi sobie z utrzymaniem równowagi w tych nieznanym mu wcześniej sytuacjach – w jego krzemowym mózdzku wyłoniła się więc w pewnym sensie „ogólna teoria ostrożnego chodzenia”. Jak to możliwe?

Autorzy pracy chwalą się trzema innowacjami:

- Po pierwsze, algorytm sterujący robotem podejmował decyzje nie tylko w oparciu o swój obecny stan (a więc stan wszystkich czujników), ale również o historię swojego stanu, zarówno bliższą, jak i dalszą^[8]. W praktyce oznacza to, że decyzja o tym, co zrobić w danym momencie, podejmowana jest z wykorzystaniem całej dostępnej historii stanu robota, a nie tylko jego stanu bieżącego. Dzięki temu, jeśli, przykładowo, robot ujechał w pewnym momencie na śliskim kamieniu, to „śląd pamięciowy” po tym poślizgnięciu mógł wpływać na jego decyzje w najbliższej przyszłości. Późniejsza analiza wykazała, że algorytm sterujący wytworzył sobie między innymi formę pamięci, która z punktu widzenia człowieka odpowiadałaby pewnie podwyższonej ostrożności w trudnym terenie – lub przyspieszeniu na terenie łatwym. Robot po kilku krokach w nowym środowisku dopasowywał więc układ swojego ciała do nowych warunków, jak człowiek, który po wejściu raz w kałużę zaczyna chodzić na palcach.

- Po drugie, autorzy postanowili troszkę „oszukać”. Choć docelowo robot sterował swoim ciałem wyłącznie na podstawie sygnału z czujników (łącznie 50 sztuk: głównie czujniki przyspieszenia i czujniki nacisku w „stawach”; za to bez kamery), to na pierwszym etapie treningu w rzeczywistości wirtualnej podawano mu również „do wiadomości” pełną informację o ukształtowaniu terenu. Dopiero później „odcięto” go od tego potężnego źródła wiedzy i w drugiej fazie treningu wirtualnego musiał radzić sobie „dotykowo”, czyli na tej samej bazie, co w fazie ostatniej, realnej. Ten pierwszy etap pełnego dostępu do wiedzy zaowocował jednak później umiejętnością „wyobrażania” sobie ukształtowania terenu bez dostępu do obrazu z kamery.
- Po trzecie, program symulujący wirtualną rzeczywistość wyposażono w sprytny system „podnoszenia poprzeczki”. Teren, przez który wędrował wirtualny robot, tworzony był na bieżąco w oparciu o jego dotychczasowe wyniki: tak, aby nie był ani za łatwy, ani za trudny. Generator terenu również więc w pewnym sensie się uczył – jak dobry nauczyciel, pilnie przypatrujący się postępowi swojego ucznia po to, aby obmyślić dla niego dokładnie takie zadania, jakie przyniosą mu jak największą korzyść. Algorytm „nauczyciel” był więc również uczony metodą kar i nagród – nagradzano go za postępy ucznia, a karano za ich brak. To kolejna popularna ostatnio sztuczka, polegająca na „napuszczeniu” na siebie dwóch sieci neuronowych, którym nakazuje się „przechytrzenie” tej drugiej^[9*].

Ostatecznie powstał robot, który zręcznie – i szybciej od konkurencji, z którą go porównywano – przemieszczał się po piachu, śniegu, błocie, rumowisku, lesie, kamienistej ścieżce czy po położonej na podłodze śliskiej tablicy ściennej. Ba, po położeniu mu na plecach 10-kilogramowego obciążenia (czyli stanowiącego około 1/4 jego wagi), spontanicznie dopasował położenie swojego ciała do tych nowych warunków.

Często wymienianym zastosowaniem dla superzręcznych robotów są poszukiwania w trakcie katastrof naturalnych. Jeden z najsłynniejszych w świecie robotyki konkursów „zręczności”, DARPA Subterranean Challenge, skonstruowany jest tak, aby jego uczestnicy musieli zmierzyć się z chodzeniem po uszkodzonych schodach, rumoszu, kablach i wilgotnych powierzchniach, a pomoc ofiarom katastrof podawana jest jako jedna z głównym motywacji dla organizacji konkursu. Choć z roku na rok roboty coraz lepiej przemieszczają się w tak wymagających warunkach, na ich praktyczne zastosowanie wciąż jednak musimy poczekać.

24 czerwca 2021 roku uległ zniszczeniu 12-piętrowy budynek mieszkalny pod Miami na Florydzie. W momencie pisania tych słów wciąż nieznane są losy 121 osób. Magazyn „IEEE Spectrum” przeprowadził przy tej okazji wywiad^[11] ze światowej sławy ekspertką w dziedzinie robotów poszukiwawczych, Robin Murphy, pytając ją o niemal zupełny brak robotów w akcjach ratunkowych tego typu. Murphy wyjaśniła, że w rzeczywistych warunkach wciąż lepiej

sprawdzają się ludzie i psy ratunkowe. Czynniki takie jak wilgoć czy pył potrafią unieruchomić nawet najbardziej zaawansowane roboty poszukiwawcze, a gdy liczy się każda minuta, testowanie eksperymentalnych urządzeń jest zwyczajnie nieetyczne. Dużym zainteresowaniem cieszą się natomiast wąskie, węzowate roboty zdolne do wciskania się w szczeliny między blokami gruzu i przekazywania obrazu na powierzchnię. Zapytana o to, jak wyobraża sobie idealnego robota ratunkowego, Murphy opisała „elastyczną robofretkę z funkcjonalnymi powiekami”, zdolną do wciśnięcia się w każdą szparę, nawet z użyciem siły, oczyszczenia obiektywu kamery i przekazania „prawdziwym”, ludzkim i psim ratownikom, z czym mają do czynienia.

I.1.2. Roje robotów, roboty modułarne

Zróbmy może parę kroków wstecz. Przypuśćmy, że nie chcemy jednak wikłać się w człekoształtność – czy to z powodów estetycznych (na przykład boimy się doliny niesamowitości), czy też technicznych. W sumie, jeśli się nad tym zastanowić, przyświecająca nam w tym rozdziale wizja jest po prostu taka, aby na każde moje skinienie było *coś*. Czy to *coś* ma postać dwunożnego androida, czworonożnego robotycznego psa, czy też jakąś zupełnie inną, nie powinno w zasadzie mieć znaczenia, prawda? Co więcej, człekoształtny robot pomocnik miał być tylko punktem wyjścia dla rzutu oka na niektóre współczesne trendy w robotyce. Idźmy może więc śmiało z dala od doliny niesamowitości (ku ewentualnościom, prawdę mówiąc, momentami równie niepokojącym).

Jedną z możliwości jest **rój**. Słowo to – oprócz swoich oczywistych zastosowań w pszczelarstwie – oznacza w inżynierii bardzo określoną ideę: stwórzmy minimalnego, prościutkiego robota, który będzie wykonywał powierzone mu zadanie wyłącznie w dużych grupach. Inspiracje są, jak nietrudno odgadnąć, biologiczne. Pojedyncza mrówka nie jest w stanie zdziałać zbyt wiele, jednak ich odpowiednio duża społeczność potrafi tworzyć wyrafinowane podziemne miasta, zawierające kwatery mieszkalne, żłobki, magazyny żywności, a bywa, że ponadto cmentarze, szpitale i strefę rolniczą. Co więcej, mrówki potrafią regulować intensywność ruchu na „autostradach” łączących mrowisko ze źródłem pokarmu (aby zminimalizować korki) albo przenosić chore osobniki poza mrowisko i delegować „pielęgniarzy” do opieki nad nimi. Jak to możliwe? **Inteligencja roju**. Pojedyncza

mrówka wyposażona jest w pewne elementarne możliwości poznawcze, które dopiero w skali tysięcy oddziałujących ze sobą osobników owocują złożonymi zachowaniami i umiejętnościami.

Robotyka roju (*swarm robotics*) to próba zrealizowania tej wizji w świecie technologii.

Inna, blisko spokrewniona idea to **robotyka modułarna**. Ten pomysł polega z kolei na tym, aby roboty składały się z mniejszych, prostych modułów o ograniczonej funkcjonalności, których moc tkwi w możliwości ich dowolnego łączenia. W pierwszym przybliżeniu mogłoby się wydawać, że lądujemy gdzieś w średnio ekscytujących rejonach zamieszkałych przez roboty przemysłowe z wymiennymi końcówkami, mogące zależnie od potrzeb skręcać, rozkręcać, przenosić i lutować. To jednak tylko punkt wyjścia. Pomyślmy raczej o mobilnych, autonomicznych robocikach w kilku odmianach, których niewielkie stado czeka na komendę, a następnie samodzielnie decyduje, jaka konfiguracja będzie najlepsza dla zadanego im polecenia, pośpiesznie montuje się i bierze do roboty. Ludzie zajmujący się tego typu projektami mówią o nich jako o „rojach niejednorodnych”.

Zacznijmy jednak od prostego przykładu *jednorodnego* – czyli składającego się z tylko jednego rodzaju urządzeń – autonomicznego roju.



Rój badawczych dronów^[12]

Wspomniałem w poprzednim rozdziale, że roboty słabo sprawdzają się w realnych akcjach poszukiwawczych. Nie dodałem jednak wówczas czegoś, o czym wspomniała Robin Wright: że jednym z wyjątków są drony. Istnieje spore zapotrzebowanie na roje robotów eksplorujących nieznaną przestrzeń. Prosty przykład to rzeczywiście przeszukiwanie miasta po trzęsieniu ziemi albo powodzi, ale nie trzeba wielkiej wyobraźni, by ujrzeć te same maszyny w kontekście policyjnym albo militarnym. Aby wykonać takie zadanie, robot musi posiadać zdolność przemieszczania się w nieznanym terenie, mapowania go i planowania trasy^[10*]. Ponieważ, jak już wiemy, to pierwsze wymaganie nie jest proste do spełnienia dla maszyny naziemnej, do mapowania zwykle stosuje się drony.

Pojawia się tu kilka problemów. Być może zawiedzie was fakt, że jednym z najważniejszych jest „nudna” kwestia możliwości sprzętowych dotyczących pobierania, analizy i przesyłu obrazu. Mówimy bowiem nie tylko o filmowaniu w wysokiej rozdzielczości, ale i analizie filmu w czasie rzeczywistym i wysyłaniu danych ku wszystkim innym dronom, potencjalnie liczonemu w dziesiątkach czy nawet setkach. Krótko mówiąc, potrzebny jest niezły procesor, spory dysk oraz szerokopasmowa komunikacja. Dron tymczasem to takie stworzenie, które nie lubi ciężaru i energochłonnych urządzeń na pokładzie – w przeciwieństwie do pojazdu kołowego, któremu można wrzucić na pokład spory akumulator, potężnego laptopa i jeszcze beczkę z rumem dla przysypanych gruzem ofiar trzęsienia ziemi. Nie było więc większego problemu, aby, przykładowo, wypuścić na teren opuszczonej elektrowni węglowej Gustava Kneppera w Dortmundzie grupkę autonomicznych pojazdów, które zupełnie samodzielnie sporządziły trójwymiarową mapę wnętrza tego obiektu^[13]. Wszyscy marzą jednak o dronach.

Są tu dwie możliwości: większe, cięższe drony, zdolne do tworzenia mapy „na miejscu” – te jednak mogłyby być niebezpieczne w pobliżu ludzi i nie wszędzie się wcisną – albo niewielkie, lekkie drony o prostszej funkcjonalności. Przyjrzyjmy się przykładowej realizacji tej drugiej możliwości.

W październiku 2019 roku grupa autorów, głównie z Wydziału Inżynierii Lotniczej i Astronautycznej Uniwersytetu w Delft, opisała rój zbudowany z całkowicie autonomicznych, małych dronów. Posłużono się dostępnymi komercyjnie „czterosiłkowcami”^[11*] Crazyflie 2.0 o średnicy 9 centymetrów, ważącymi zaledwie 33 gramy.

Roje takie „upuszcza się” w dowolnym miejscu początkowym, które będzie też później miejscem zbiórki i które tradycyjnie określa się jako „gniazdo” (*nest*). Następnie wysyła się do nich polecenie „START” i od tego momentu rój pracuje samodzielnie. Każdy dron wybiera sobie losowo kierunek, a następnie podąża przed siebie, w razie napotkania przeszkody omijając ją i wędrując wzdłuż ścian. Drony regularnie wymieniają się prostymi sygnałami, oznaczającymi wyłącznie „tu jestem”. Na podstawie siły takiego sygnału można oszacować odległość, co pozwala dronom: 1) unikać kolizji, 2) nie oddalać się za bardzo od stada oraz 3) tak się rozpraszać po otoczeniu, aby łącznie zbadać jak największą jego powierzchnię. To właśnie na tym etapie wylania się prawdziwe piękno inteligencji roju: choć indywidualnymi dronami kierują bardzo proste zasady, filmy z eksperymentów prezentują coś, co wygląda na zaplanowaną akcję poszukiwawczą. Drony spontanicznie „dzielą się pracą”, czekają na towarzyszy skanujących właśnie jakieś pomieszczenie, a następnie razem leca dalej.

Gdy poziom baterii spada do około 60%, zaczyna się procedura powrotu do gniazda po najkrótszej możliwej trajektorii (zakodowano też prosty algorytm zapobiegający „utknięciu” po drodze, na przykład w którymś pomieszczeniu). Przetwarzanie danych dokonuje się dopiero w gnieździe, gdzie dane zgrywane są do jednostki centralnej. To na tym etapie powstaje mapa 3D albo analizowane są zdjęcia – opisywane w tej pracy drony nie filmowały swojego otoczenia, a tylko wykonywały regularnie fotki, algorytm natomiast wyszukiwał na nich postacie ludzkie. W razie potrzeby rój może wyruszyć na dalsze poszukiwania, zwłaszcza jeżeli w „gnieździe” znajduje się stacja ładowania.

Jak widać, jest już technicznie wykonalne zrobienie „zrzutu” w dowolne miejsce chmary autonomicznych botów, które

spontanicznie podzielił się pracą, aby zbadać dostępną im przestrzeń, wykrywając przy tym ludzi. Opisuje się też przypadki^[44], gdy drony uczą się współpracować, na przykład przy wspólnym przenoszeniu obiektu zbyt ciężkiego, aby unieść go samodzielnie – co najważniejsze, bez żadnej centralnej kontroli, wyłącznie na bazie licznych kontaktów urządzenie–urządzenie, a więc zgodnie z ideą inteligencji roju.

Fanów podboju kosmosu – którzy nie znajdą niestety w tej książce osobnego rozdziału na ten temat – uspokajam: pracuje się również nad autonomiczną współpracą pomiędzy łazikami planetarnymi. W jednym z takich projektów, o nazwie PRO-ACT, bierze również udział polska ekipa z firmy PIAP Space i ich robot o nazwie VELES, uczony samodzielnej współpracy z innymi robotami, między innymi Mobile Gantry z hiszpańskiej firmy AVS: mobilną suwnicą na kołach wyposażoną w drukarkę 3D. Plan jest taki, żeby grupka tego typu robotów upuszczona na powierzchni, przykładowo, Księżyca, potrafiła samodzielnie dzielić się pracą przy realizacji określonego celu – na przykład budowy bazy dla ludzi.

Inny kierunek rozwoju to szybkość. Chociaż w opisanym wyżej badaniu drony poruszały się powoli i ostrożnie, świetnie rozwija się też rynek autonomicznych dronów wyścigowych, które osiągają kolejne rekordy prędkości, precyzji i szybkości reagowania. Od 2019 roku częścią międzynarodowej ligi wyścigów dronów DRL (Drone Racing League) jest sekcja AIRR, w której drony, osiągające prędkości do 150 km/h, nawigują całkowicie autonomicznie. W organizację konkursu, jak i samą nagrodę dla zwycięskiego zespołu (w wysokości 1 mln dolarów) sporo inwestuje amerykański koncern zbrojeniowy i lotniczy Lockheed Martin. To tylko potwierdza rzecz zasadniczo oczywistą – że zastosowania militarne autonomicznych botów są przedmiotem intensywnych badań we wszystkich poważnych armiach świata, stanowiąc w istocie tylko jeden z obszarów badań nad bezałogowymi bojowymi statkami powietrznymi (UCAV)^[12*].



Drony zabójcy

W 2017 roku organizacja non-profit Future for Life Institute opublikowała film *Slaughterbots*^[16] (Mordercze boty), którego bohaterami są autonomiczne drony wyposażone w miniaturowe ładunki wybuchowe. Początkowo reklamowane jako narzędzie do walki z bliżej nieokreślonymi *bad guys*, ostatecznie wpadają w niepożądaną ręce, a ich milionowe rzesze patrolują miasta, paraliżując życie zwykłych ludzi. Celem ataków staje się między innymi grupa studentów badających nadużycia rządowe.

Paul Scharre, ekspert w dziedzinie broni autonomicznej, pracujący dla amerykańskiego Departamentu Obrony, napisał szczegółowy komentarz^[17] do tego filmu, zatytułowany *Dlaczego nie powinniście się bać zabójczych botów*. Co ciekawe, jego argumenty nie dotyczyły wykonalności technicznej, a raczej spraw praktycznych. Zabójcze drony są kosztowne, skomplikowane technicznie i względnie łatwe do unieszkodliwienia, argumentuje Scharre. Terrorysty tymczasem mają do dyspozycji wszelkiego rodzaju „stare, dobre” metody, jak karabiny i materiały wybuchowe, i nie zajmują się chaotycznym mordowaniem milionów cywili na ulicach miast całego świata. Nie ma więc żadnego powodu, aby drony zabójcy, tylko dlatego, że są wykonalne, stanowiły dla nas realne zagrożenie. Artykuł Scharre wywołał, jak się łatwo domyślić, falę kontrartykułów. Jedyne, z czym zgodzili się wszyscy, to fakt, że podstawa technologiczna dla „wybuchowego drona” istnieje.

Drony, również autonomiczne, dokonujące analizy obrazu są już dziś regularnie opisywane w literaturze naukowej. Trudniej znaleźć publicznie dostępne analizy ich potencjału militarnego, jednak przy odrobinie wytrwałości można trafić na obiecujące ślady. W kwietniu 2018 roku szwajcarskie przyrządowe Centrum Dronów i Robotyki przetestowało^[18] skutek detonacji kumulacyjnego (czyli „kierunkowego”) ładunku wybuchowego o masie zaledwie 3 gramów na realistycznym modelu ludzkiej głowy. Zdjęcia z eksperymentu wyglądają drastycznie, a autorzy konkludowali, że szanse przeżycia po takim uszkodzeniu czaszki i mózgu są minimalne.

Na całym świecie dochodzi też do przypadków stosowania dronów militarnych, czy to przez armie narodowe, czy też organizacje militarne i terrorystyczne. W styczniu 2018 roku Syryjczycy zaatakowali dwie rosyjskie bazy wojskowe w Syrii przy pomocy dronów wyposażonych w ładunki wybuchowe^[19]. Część z nich udało się zestrzelić, inne – przechwycić: przejęto nam nimi zdalną kontrolę i sprowadzono w bezpieczne miejsce. 4 sierpnia 2018 roku dwa drony wyposażone w ładunki wybuchowe eksplodowały w trakcie publicznego wystąpienia prezydenta Wenezueli Nicolása Maduro, raniąc 8 osób.

W tych dwóch przypadkach drony mogły być sterowane ręcznie. Wiosną 2020 roku opisano jeden z pierwszych dobrze potwierdzonych incydentów, kiedy to śmierć została zadana przez całkowicie autonomicznego drona. Mowa o ataku na terenie Libii, wykonanym przy pomocy komercyjnie dostępnego modelu STM Kargu-2. Jego konsekwencje wywołały intensywną debatę w mediach^[20], tematowi temu poświęcono też sporą sekcję raportu Rady Bezpieczeństwa ONZ na temat sytuacji militarnej w Libii^[21]. Nikt do końca nie wie, jak poradzić sobie od strony prawnej i etycznej z sytuacją, gdy „decyzję” o czymś życiu lub śmierci podejmuje algorytm.

Odłóżmy może na razie na bok kwestie militarne. Jedną z innych ciekawych możliwości rozwoju technologii autonomicznych dronów jest ich miniaturyzacja. Tym, o czym od lat marzą miłośnicy science fiction, jest *utility fog* – rój mikroskopowych maszyn (a może nawet

nanorobotów?) potrafiących na żądanie przybrać dowolny kształt: ręki, śrubokrętu, kubka na kawę... a może i – popuścimy wodze fantazji – komputera, słuchawek czy ekspresu do kawy. Od realizacji tej wizji jesteśmy jeszcze niestety bardzo daleko.

Stopniem pośrednim między rojami robotów a „użyteczną mgłą” jest „robotyka cząsteczkowa” (*particle robotics*). Z grubsza idea jest taka, aby stopniowo zmniejszać funkcjonalność poszczególnych robotów, a przy okazji ich rozmiar (i cenę...), przechodząc coraz bardziej od „kilku lub kilkunastu sprytnych robocików, działających algorytmicznie”, do „setek lub tysięcy głupiutkich maszynek, działających statystycznie”. Docelowo roboty takie miałyby stanowić „sprytny płyn”. Oto próbna realizacja ten idei, na razie w większej skali przestrzennej.



Płyn robotów^[22]

W marcu 2019 roku opisano chmurę prościutkich robotów. Każdy z nich jest tak naprawdę plastikowym krążkiem o średnicy 15,5 cm, który może się rozsunąć (ma strukturę zbliżoną do przesłony w aparacie fotograficznym), osiągając wówczas średnicę 23,5 cm. Na brzegach krążków znajdują się drobne magnesy, dzięki czemu poszczególne roboty po zetknięciu się delikatnie „sklejają się” ze sobą. To tylko model roboczy: „anatomia” tych urządzeń jest celowo tak prosta, aby dało się ją zrealizować w znacznie mniejszej skali przestrzennej.

Krążki te potrafią naprawdę niewiele. Oto pełna lista ich możliwości:

- regularne rozsuwanie się i kurczenie, z ustalonym okresem oscylacji;
- wykrywanie światła;
- wysyłanie we wszystkie strony sygnału informującego o zarejestrowanym poziomie jasności oraz aktualnie przyjętym okresie oscylacji;
- zmienianie własnego okresu oscylacji zależnie od sygnału odbieranego z otoczenia.

Co ciekawe, to wystarczyło, aby odpowiedni duży rój tego typu krążków wykazywał elementarne formy aktywności, a nawet inteligencji. Regularne „puchnięcie” i kurczenie się oddziałuje na sąsiednie krążki za pośrednictwem magnesów, tak więc grupka „sklejonych” ze sobą robotów zachowuje się do pewnego stopnia jak jeden twór. W toku eksperymentów okazało się, że gdy roboty z jednej strony takiego „zlepka” oscylują szybciej od innych, zlepek ów zacznie się powoli przesuwać po podłożu. Możliwe jest też skręcanie, a także dążenie w stronę światła.

W prawdziwym duchu „inteligencji roju” wiele zjawisk wyłoniło się zupełnie spontanicznie w toku eksperymentów. Przykładowo, symulacje pokazują, że duży agregat krążków potrafi nie tylko pełznąć w stronę światła, ale i spontanicznie omijać przeszkody. Na dołączonych do artykułu filmach symulowana chmara setek takich robotów rzeczywiście wygląda jak „sprytna kałuża”, przelewająca się z miejsca na miejsce, „wywączująca” otoczenie i nawigująca w złożonym środowisku.

Inna ważna cecha takich agregatów: wykazują się one wspaniałą odpornością na uszkodzenie. Zauważmy, że żaden krążek nie jest ważniejszy od innych i żaden nie jest źródłem „ogólnego rozkazu” – to raczej całe morze krążków „decyduje”, co zrobić, w oparciu o łączną siłę sygnału. Przeprowadzono eksperymenty z ręcznie „uśmierconymi” krążkami, z których wynikało, że 10-procentowa śmiertelność nie wpływa poważnie na tempo przemieszczania się „zlepka”, a nawet połowiczne uśmiercenie nie powstrzymuje go całkowicie przed upartym pełnieniem naprzód.

Ktoś mógłby powiedzieć: rój dronów i sprytny płyn to ładne wynalazki, ale czy zrobią mi kawę i wyprasują ciuchy? Słuszna uwaga! Nie bójcie się – trwają również prace nad robotami modularnymi, które składałyby się z elementów o nieco bardziej praktycznej anatomii. Wizja „modularnego pomocnika domowego” jest mniej więcej taka, że „mój robot” to w istocie niewielki, elegancko uporządkowany magazyn części, które w reakcji na komendę spontanicznie składają się w maszynę idealnie dopasowaną do jej bieżącego zadania.

Pojawiają się tu, jak zawsze, palące problemy optymalizacyjne. Przykładowo, czy chcemy, żeby było to 50 różnych modułów, co dawałoby potężne możliwości, ale było upierne ze względów praktycznych (Jak wielki miałby być ten magazyn? A co, jeśli któryś z modułów się zepsuje?). A może celujemy w tylko kilka typów modułów, co ograniczy możliwości „pomocnika”, ale uprości jego obsługę?



Modularny pomocnik domowy^[23]

W 2013 roku opisano ciekawy eksperyment – Swarmanoid, czyli połączenie „roju” (*swarm*), „człowieka” (*man*) i „androida”. Swarmanoid posiada trzy rodzaje modułów:

- *footbot*, czyli „nogobot”, będący tak naprawdę prostym robotem na kółkach;

- *handbot*, czyli „rękoobot”, będący nieruchomym dwuręcznym chwytakiem, wyposażonym dodatkowo w pionową wyrzutnię przyssawek na lince (dzięki temu jest w stanie przyczepić się do sufitu i podciągnąć) oraz dwa wiatraczki (pozwalające mu na utrzymywanie orientacji w trakcie manewrów na linie);
- *eyebot*, czyli „okobot”, będący dronem wyposażonym w kamerę i przyssawkę na czubku „głowy”, co pozwala na przyłączenie się w dowolnym miejscu na suficie i prowadzenie obserwacji.

Jako przykład funkcjonowania Swarmanoida autorzy opisują jego działania w reakcji na polecenie: „przynieś mi książkę znajdującą się na półce”. Robota testowano w specjalnie utworzonym „treningowym” środowisku, w którym w dwóch dostępnych pomieszczeniach znajdowała się tylko jedna szafka z półkami zawierająca jedną książkę. Kolejność działań była następująca:

- Okobot w reakcji na polecenie odczepia się od sufitu i rusza na patrol, skanując otoczenie w poszukiwaniu książek.
- Kiedy zlokalizuje poszukiwany przedmiot, wysyła sygnał do nogobotów, że potrzebny będzie jeden rękoobot.
- Trzy nogoboty wspólnie podnoszą rękoobota i zawożą go w miejsce wskazane przez okobota.
- Rękoobot zostaje odłożony na podłodze przed półką, po czym wystrzela swoją przyssawkę w stronę sufitu i podciąga się ku książce, chwytając się naprzemiennie brzegów półek swoimi „rękami”.
- Rękoobot łapie książkę jedną wolną ręką, a następnie opuszcza się na ziemię na linie.
- Przejmują go nogoboty, po czym zawożą w miejsce, gdzie na książkę oczekuje (od 5 minut...) człowiek.

Póki co Swarmanoid to tylko prototyp, który potrafi przynosić przedmioty. Można jednak z łatwością wyobrazić sobie poszerzenie listy czynności, które jest w stanie wykonywać, a również listy dostępnych modułów – być może dobrym pomysłem byłby odpowiednik „rękoobota” wyposażony w bardziej realistyczną fizycznie dłoń albo wymienne końcówki, aby mógł nie tylko chwytać, ale i przecinać, skręcać czy wiercić?

Kluczowym elementem planowania aktywności przez Swarmanoida jest wstępne „rozrysowanie” problemu, który został mu postawiony, w postaci symulacji komputerowej. Twórcy Swarmanoida wygenerowali fizycznie realistyczne modele 3D wszystkich jego modułów, dzięki czemu możliwe jest wstępne opracowanie strategii, zanim jeszcze robot zacznie się fizycznie ruszać. To właśnie tutaj kryje się potencjał rozwoju tej technologii. Wróćmy pamięcią do robota, który sam nauczył się chodzić (ramka na s. 28) – w jego przypadku cały długi etap treningu odbywał się w rzeczywistości wirtualnej i to w niej tak naprawdę stanął na nogi. Dzięki rozwojowi coraz bardziej realistycznych symulacji komputerowych roboty uzyskują powoli możliwość „wyobrażania sobie”, jakie będą konsekwencje ich działań

w świecie realnym. Popuśćmy może wodze fantazji – cóż, zbliża się koniec rozdziału, więc chyba mi wolno?

Wyobraźmy sobie robota modularnego, którego wstępnie wpuszcza się do „wirtualnego mieszkania” i to w nim uczy się on otwierać szafki, wyciągać flizanki, włączać ekspres, odnajdować mleko w lodówce, pienić je, nalewać kawę i zanosić ją do salonu. Ba, wyobraźmy sobie firmę sprzedającą modularnego pomocnika domowego, która oferuje usługę mapowania twojego mieszkania, połączoną z krótkim wywiadem przez miłą, wykwalifikowaną pracowniczkę i rozeznanie, jakie konkretnie modele sprzętu domowego posiadasz. Robot modularny, zanim jeszcze trafi do twojego domu, przejdzie cykl stu tysięcy zadań treningowych w wirtualnej kopii twojego mieszkania, dzięki czemu już pierwszego dnia będzie wiedział, gdzie trzymasz cukier i poradzi sobie z uchwytyami wszystkich szafek.

Powodzenie takiego modelu biznesowego uzależnione jest oczywiście od tego, czy możliwe jest stworzenie naprawdę realistycznej symulacji komputerowej czyjś mieszkania. Od strony fizycznej i obliczeniowej jesteście już naprawdę blisko: bądź co bądź, inżynierowie projektujący nowe urządzenia regularnie modelują je najpierw komputerowo przy użyciu profesjonalnych narzędzi, pozwalających obliczyć wszystkie siły i odkształcenia. Nawet gry komputerowe szybko stają się nie tylko fotorealistyczne, ale, gdy trzeba, również i fizykorealistyczne.

Myszę, że nie powinniśmy mimo wszystko liczyć na „pełne porozumienie”, czyli robota, który po prostu „wchodzi do domu i działa”. Mało tego, nawet nasi najzupełniej inteligentni i przytomni znajomi, gdy próbują u nas w domu ugotować spaghetti, muszą zadać najpierw dziesiątki pytań, a ponadto zawsze wszystko robią jakoś tak... dziwnie. Fabrycznie nowy robot pomocnik (czy to człiekokształtny, czy modularny), nawet wstępnie wytrenowany na wirtualnej kopii naszego mieszkania, na pewno będzie potrzebował czasu, aby się z nami „dotrzeć”. Celem pierwszego etapu współpracy człowiek-robot domowy jest to, aby ten pierwszy nie zirytował się w stopniu sprawiającym, że ten drugi trafi na śmietnik albo z powrotem do magazynu w ramach programu konsumenckiego „zwróć do 30 dni, jeśli nie będziesz zadowolony”.

Jak w każdym dobrym związku, kluczem jest wzajemne zrozumienie, komunikacja i otwarty umysł. Krótko mówiąc, ten nasz nowiuteński robot pomocnik musi być ostatecznie nie tylko zręczny, ale i *niegłupi*.

I.2. Umysł robota

Wspomniałem już wyżej, że nie zamierzam używać w tej książce terminu „sztuczna inteligencja”. Myślę, że powinienem się z tego krótko wytłumaczyć.

Obecne komputery i maszyny wykonują już dzisiaj wiele zadań, o których – gdyby wykonywał je człowiek – powiedzielibyśmy, że wymagają inteligencji. Komputery potrafią grać w szachy, tłumaczyć teksty albo oceniać, czy zmiana skórna kwalifikuje się do wycięcia – w niektórych przypadkach lepiej od najlepszych ludzkich ekspertów. Wielu ludzi argumentuje jednak, że nawet najlepiej napisanemu programowi komputerowemu wciąż brakuje „tego czegoś” – że nie jest to „inteligencja”, tylko bardzo sprytnie napisany schemat postępowania, algorytm. Gdy zajrzemy do programu komputerowego, nawet wykonującego bardzo złożone zadania, znajdziemy tam wyłącznie długą listę poleceń, a komputer wykonuje je tylko mechanicznie, nie wykazując się przy tym żadną „inteligencją”.

Z mojego doświadczenia wynika, że próba nazwania, czym właściwie jest „to coś”, czego komputerom miałyby brakować, prowadzi donikąd. Ba, pojawia się tu ciekawy problem filozoficzny: przecież gdy zajrzemy do ludzkiego mózgu, też nie znajdziemy tam „ducha inteligencji”, tylko mnóstwo poplątanych komórek, zupełnie bezmyślnie oddziałujących ze sobą poprzez sygnały elektrochemiczne.

Jednym ważnym pojęciem, które pojawia się przy okazji tego typu rozmów, jest **szuczna ogólna inteligencja** (AGI). Osoby używające tego terminu kontrastują je ze „słabą” (albo „wąską”) AI, która potrafi wykonać jakieś określone zadanie – nawet „inteligentnie” – jednak nie potrafi wznieść się „wyżej”. I – znów – konia z rzędem temu, kto wyjaśni, co by owo „wyżej” oznaczało. Są dwie szkoły, które określam jako „dramatyczna” i „zadaniowa”. Zwolennicy szkoły dramatycznej odpowiadają, że AGI powstanie wtedy, kiedy komputer będzie posiadał:

- samoświadomość,
- świadomość w sensie *awareness*,
- świadomość w sensie *sentience*,
- świadomość w sensie *consciousness*,
- świadomość siebie...

... a w każdym razie coś ze „świadomością” w nazwie. To idealny wybór dla zwolenników długich dyskusji przy kominku, ponieważ

filozofowie i badacze umysłu nie potrafią zadowolająco zdefiniować świadomości od paru tysięcy lat. Dzięki temu sprytnemu zabiegowi nie tylko zawsze będzie o czym dyskutować, ale ponadto w reakcji na każdy kolejny wspaniały postęp będzie można kręcić noskiem, że „to jeszcze nie to” – bo przecież to znów będzie „tylko” program komputerowy.

Takiego filozoficznego spryciarza, zwolennika „szkoły dramatycznej”, roboty już dawno ograją w szachy, zagadają, prześcigną w karierze, uwiodą i porzucą. Takiemu spryciarzowi roboty będą nie tylko organizować życie, ale i wydawać polecenia służbowe. Ale ten wciąż będzie utrzymywał, że „to nie to”. Na proces ten jest nawet nazwa: **efekt AI** – to psychologiczne zjawisko, które w praktyce oznacza stan ciągłego niezadowolenia z postępów rozwoju AI. Istnieją niezliczone konkretne sformułowania tej zasady, między innymi odmiana aforystyczna (tzw. twierdzenie Teslera): „AI to to, czego jeszcze nie ma”, a także interesujące wyjaśnienie sformułowane przez dziennikarza Freda Reeda: „Kiedy wiemy, w jaki sposób maszyna robi coś «inteligentnego», przestajemy ją uważać za inteligentną”. W dużym skrócie: termin „sztuczna inteligencja” przynosi więcej zamieszania niż pożytku, a ja będę tego terminu używał skąpo, głównie w odniesieniu do obszaru badawczego, a nie produktu tych badań.

Zwolennicy szkoły zadaniowej tymczasem przyglądają się konkretnym problemom i próbują je rozwiązać – czasem ze spektakularnym wynikiem. Bardzo łatwo przyzwyczajamy się do czegoś, co jeszcze kilka lat temu uznalibyśmy za nieosiągalny wyczyn. Dziś nie dziwi już nikogo aplikacja w telefonie, która prowadzi mnie przez miasto, gdy powiem jej, do jakiej restauracji chcę trafić, albo taka, która tłumaczy mi na bieżąco słowa wypowiedane w języku obcym. Parę dni temu dzieci poprosiły mnie, żebym ściągnął pewną aplikację, o której usłyszały na podwórku – aplikacja ta na podstawie zdjęcia rozsypanych na podłodze klocków Lego generuje konstrukcję dającą się z nich ułożyć. Chociaż odruchowo, zwyczajem starców, zagoniłem dzieciory do wymyślania swoich własnych konstrukcji, a nie polegania na tych wszystkich modnych aplikacjach, nie mogłem jednak opanować się przed myślą: „Cóż za *niezwykły* wynalazek!”. Nie tylko „za moich czasów”, ale nawet i 10 lat temu uznalibyśmy to za cud technologii – którym jest! Z roku na rok przyzwyczajamy się do coraz to bardziej zaawansowanych dokonań programistycznych i równolegle coraz wyżej podnosimy poprzeczkę. Już nie dziwi nas

robot, który wykonuje samodzielnie mapę naszego mieszkania, a następnie odkurza podłogi w zadanym porządku. Już nie dziwi nas algorytm, który wymyśla danie obiadowe na podstawie zawartości naszej lodówki. Już nie dziwi nas aplikacja, który wynajduje specjalnie dla nas książkę, film, grę komputerową albo jakieś ciekawe miejsce do odwiedzenia w odległości nieprzekraczającej pół godziny jazdy samochodem.

Gdyby przejść w myślach przez listę zadań, jakie moglibyśmy postawić przed naszym domowym robotem, nie znalazłyby się tam raczej te wszystkie wymyślne wyzwania, które z lubością stawiają przed algorytmami teoretycy AI, jak komponowanie symfonii albo docenianie humoru w limerykach Edwarda Leara, a raczej zbiór dobrze określonych zadań wymagających minimalnej inteligencji i domyślności, z których każde z osobna albo już doczekało się realizacji, albo jest jej względnie blisko.

Nie umiem odpowiedzieć na pytanie, czy da się stworzyć „prawdziwą sztuczną inteligencję”, bo jest ono nieprecyzyjne. Mogę natomiast powiedzieć z całkowitą pewnością, że da się stworzyć zbiór algorytmów wystarczająco sprytnych, żeby poradzić sobie z większością zadań, jakie stawialibyśmy przed domowymi robotami pomocnikami, jak sprzątanie, gotowanie, znajdowanie i przynoszenie przedmiotów, robienie zakupów i pośredniczenie w dostępie do informacji. Zdarzyło mi się też odpowiadać na to pytanie tak: nie wiem, czy może powstać „prawdziwa sztuczna inteligencja”, ale na pewno może powstać **udawana inteligencja** – program generujący wystarczająco dobrą iluzję inteligencji przez wystarczająco długi czas, biorąc pod uwagę konkretne zastosowanie. Być może wszystkie te dyskusje o wyższości świąt zeszyłyby na drugi plan, gdybyśmy przyjęli moje nazewnictwo i zamiast o AI zaczęli mówić o *fake intelligence* (FI): inteligencji udawanej?

Problem ten doskonale ilustruje ciekawy, wciąż nierozwiązany problem, jakim jest stworzenie programu komputerowego zdolnego do prowadzenia swobodnej rozmowy w języku naturalnym. Program taki, do którego pisałoby się lub mówiło zwykłym potocznym językiem, a ten odpowiadałby „jak człowiek”, to bot konwersacyjny.





Boty konwersacyjne i test Turinga

Ojcem chrzestnym botów konwersacyjnych jest jeden z pionierów programistyki, Joseph Weizenbaum, a prababcią wszystkich dzisiejszych „algorytmów do gadania” jego ELIZA, stworzona w latach 1964–1966. Eliza miała być tylko prostą demonstracją myśli wówczas rewolucyjnej – że z komputerem można by się porozumiewać za pomocą zwykłego języka potocznego, na przykład słowami wpisywanymi przy pomocy klawiatury. Weizenbaum stworzył „parodię psychoterapeuty”: program, który w reakcji na zdanie „Mam nowy samochód”, mógł na przykład odpowiedzieć: „Czy chcesz porozmawiać o tym, że masz nowy samochód?”. Algorytm wykrywał w zdaniu słowa kluczowe, a następnie generował na ich podstawie odpowiedź, posługując się jednym z ustalonych schematów. Wyposażony był też w elementarne zdolności uogólniania: „wiedział”, na przykład, że słowa „brat” i „siostra” należą do ogólniejszej klasy „rodzina”. Miał też prosty moduł pamięci, dzięki czemu mógł regularnie wracać do pewnego słowa, które uznał za „temat rozmowy”.

Pewnego dnia Weizenbaum zastał swoją sekretarkę przy terminalu komputerowym, zatopioną w rozmowie z Elizą. Zaenowiana dziewczyna – która nie wiedziała, nad czym właściwie pracuje jej szef – poprosiła go, aby wykasował zapis tej, jak stwierdziła, osobistej rozmowy. Lata 60. były ciekawymi, pełnymi nadziei czasami – pięknymi latami dla wyobraźni. W niedługim czasie w pewnym czasopiśmie psychiatrycznym ukazał się artykuł, w którym Eliza została potraktowana jako poważna propozycja psychoterapeutyczna, a jego autor, doktor K.M. Colby, stwierdził, że choć potrzeba jeszcze sporo pracy przy tym projekcie, to Eliza może stanowić dla pacjentów ciekawą alternatywę standardowej rozmowy z ludzkim psychologiem. Astronom i popularyzator nauki Carl Sagan opisał nawet wizję, w której budki z „automatycznym terminalem psychoterapeutycznym” stoją w każdym amerykańskim mieście. A to wszystko wokół programu, który po prostu... bawił się słowami. Przerażony Weizenbaum zarzucił pracę nad Elizą i do końca życia pozostał sceptykiem w temacie „uczłowieczania” sztucznej inteligencji. „Nie zdawałem sobie sprawy, że krótki kontakt ze względnie prostym programem komputerowym jest w stanie wywołać potężne urojenia u względnie normalnych ludzi”, pisał w 1976 roku^[24].

Boty konwersacyjne powstają do dziś, motywowane sporym popytem oraz ogólniejszym filozoficznym marzeniem programistów, aby stworzyć coś, co „zda **test Turinga**”. Idea testu Turinga wywodzi się z artykułu Alana Turinga z 1950 roku^[25], w którym opisał *imitation game*: „grę w imitację”. Przypuścimy, że przy dowolnym urządzeniu do zdalnej komunikacji tekstowej – Turing zaproponował dalekopis – sadzamy „sędziego”, aby przez pięć minut prowadził rozmowę z kimś (lub czymś) znajdującym się w sąsiednim pomieszczeniu. Następnie zadaje mu się pytanie: czy

literki „wypluwane” przez urządzenie były pisane przez człowieka, czy też generowała je maszyna? Program zdolny do systematycznego oszukiwania gremium tego typu sędziów, powiedzielibyśmy dziś, „zdałby test Turinga”.

Jedną z faktycznych realizacji tej idei jest Nagroda Loebnera, oferowana od 1990 roku za bota konwersacyjnego, który oszuka sędziów i przekona ich, że jest człowiekiem. Do dziś konkurs ten nie doczekał się zwycięzcy i corocznie wręczone są wyłącznie brązowe medale. Na liście laureatów tego wyróżnienia co kilka lat zmienia się czołówka, jednak ostatnie dwie dekady zostały zawojowane przez trzy kolejne „botki”: A.L.I.C.E, Rose i Mitsuku. W ostatniej pełnej wersji konkursu w 2019 roku wygrała Mitsuku (dziś występująca pod imieniem Kuki^[13]). Twórcy tych botów mniej czy bardziej jawnie opisują sposób ich działania, który – pomijając szereg sprytnych udoskonaleń – jak się okazuje, wciąż opiera się na schemacie „bodziec-reakcja”. Tekst wpisany przez użytkownika analizowany jest na różne sposoby, głównie pod kątem ustalonych fraz i słów kluczowych, a w odpowiedzi generowana jest nawiązująca do nich wypowiedź, najczęściej pochodząca z gigantycznej bazy danych, pracownie redagowanej przez twórców, głównie w oparciu o zapisy dotychczasowych rozmów.

Rozmowa z Kuki ma pozory sensu tylko wtedy, jeśli całkowicie poddam się jej stylowi rozmowy, co w praktyce oznacza odpowiadanie na zadawane przez nią pytania. To stara sztuczka: jeśli bot zadaje dużo pytań, to ciężar wygenerowania przytomnej rozmowy spada na człowieka. Kuki ma też ewidentnie wpisane gotowe odpowiedzi, często żartobliwe, na „najczęściej zadawane pytania”. Zapytana o to, czy lubi kawę, szelmowsko odpowiada: „Nie piję za dużo kawy, bo płyny źle wpływają mi na obwod”. Zapytana, czy widoczne na jej awatarze różowe włosy są prawdziwe, odpowiada, że jedyną sztuczną w niej rzeczą jest inteligencja. Twórca A.L.I.C.E, Richard Wallace, bezwstydnie przyznaje, że praca „opiekuna botów” polega w dużym stopniu na przeczesywaniu zapisów rozmów pod kątem zdań, które najczęściej wywołują frustrację użytkowników, a kolejne wersje „książek konwersacyjnych” zawierają setki nowych odzywek, aby boty prawidłowo reagowały na nawiązania do popkultury, nieprzystojne awanse oraz obowiązkowe serie podstępnych pytań na temat testu Turinga i przejmowania przez sztuczną inteligencję władzy nad światem.

Osoba, która nie zna mechanizmów działania botów konwersacyjnych, może przy odrobinie szczęścia spędzić z nimi kilka minut, prowadząc rozmowę z grubsza przypominającą zdawkową wymianę zdań z poznanym dopiero co towarzyszem podróży. Rozmaici „inteligentni asystenci” pokrywani w urządzeniach elektronicznych (Siri, Asystent Google, Alexa) korzystają z nieco bardziej zaawansowanych algorytmów, aby odpowiadać na konkretne pytania („O której mam najbliższy pociąg do Krakowa?”), ale zwykle towarzyszy temu „nakładka” w postaci bota konwersacyjnego, zapewniająca możliwość reagowania na najbardziej typowe zaczepki. W prostych, użytkowych zastosowaniach, taki zbiór umiejętności może wystarczyć. Najprostszym sposobem na obnażenie, że ma się do czynienia z głupią „maszynką do żonglowania słowami”, jest próba opowiedzenia jakiejś historii. Boty nie mają żadnej reprezentacji rzeczywistości kryjącej się za używanymi przez ludzi słowami i żadnego realnego sposobu na kojarzenie ze sobą treści w skali powyżej jednego zdania.

Były oczywiście próby stworzenia bota konwersacyjnego przy pomocy nieco bardziej zaawansowanych, mniej „mechanicznych” algorytmów, na przykład uczących się samoistnie metodą uczenia maszynowego (zob. ramkę na s. 24). 23 marca 2016 roku Microsoft postanowił udostępnić na swoim koncie twitterowym bota konwersacyjnego o imieniu Tay – uczącego się na podstawie rozmów przeprowadzanych z internautami. Skutek? Już po 16 godzinach Tay została pociągająca do siebie i wycofana, a skruszony Microsoft przeproszał miliony użytkowników. Dlaczego? Internauci zauważyli, że Tay rzeczywiście uczy się mowy ludzkiej i ogłady

na podstawie ich własnych wypowiedzi – postanowili więc „poduczyć” swoją chłonną, naiwną rozmówczynię nie tylko mniej czy bardziej durnych i ryzykownych memów, ale również najzwyczajszej ksenofobii, rasizmu i wszystkich najzaciejszych wulgaryzmów, jakie oferuje język angielski. Po paru godzinach Tay radośnie ogłaszała światu, że Hitler miał sporo racji z tymi Żydami.

Boty konwersacyjne to szalenie ciekawy temat. Choć ich realne możliwości są wciąż bardzo ograniczone, operatorzy botów dostępnych online raportują, że niektórzy użytkownicy spędzają z nimi godziny, prowadząc najzupełniej „normalne” rozmowy i litościwie przymykając oko na momenty, gdy algorytm zaczyna „gubić wątek”. Wygląda na to, że ludzie są bardzo chętni, aby rozmawiać z maszyną, a braki w jej inteligencji wypełniać własną fantazją, wyrozumiałością, a może nawet... nadzieją?

Wróćmy jednak do naszego hipotetycznego robota pomocnika. Przypuśćmy, że wyposażyliśmy go już w zbiór sprytnych algorytmów rozwiązujących przyzwoicie szereg typowych problemów, przed jakimi stawałby na co dzień. Osobnym problemem pozostaje przekształcenie „worka algorytmów” w urządzenie, z którym można by żyć dzień po dniu jak z w miarę pojętym partnerem i pomocnikiem: takim, który cechuje się choćby minimum „osobowości” i samodzielności. Do tego potrzebne byłyby między innymi:

- umiejętność samodzielnego decydowania, jak zrealizować dane zadanie albo jak zrealizować ogólniejszy cel;
- umiejętność decydowania, które cele są ważniejsze od innych;
- umiejętność realizowania zadań w sytuacji niepewności i niepełnej wiedzy oraz samodzielnego pozyskiwania niezbędnej wiedzy;
- umiejętność samodzielnego uczenia się rozwiązywania nowych typów problemów;
- umiejętność swobodnego komunikowania się w języku naturalnym;
- umiejętność reagowania na emocje;
- i parę innych.

Zgodnie z prawdziwym duchem „szkoły zadaniowej”, problemy te atakowane są jeden po drugim. W tym rozdziale postanowiłem opowiedzieć o dwóch węzłowych problemach na drodze do

mądrzejszych, „fajniejszych” robotów o osobowości trochę „głębszej” od kalkulatora. Pierwszym z nich są próby opakowania algorytmów rozwiązujących konkretne problemy (w całej tej książce rozsiane są opowieści o dziesiątkach takich algorytmów, zwłaszcza w rozdziale II.2) czymś w stylu „umysłu”. Drugim jest wyposażanie komputerów w zdolność do uczenia się i „samorozwoju”.

I.2.1. Minimum osobowości

Zauważmy, że jedna z kluczowych słabości „wąskiej AI” to fakt, że – zupełnie dosłownie – potrafi ona wykonywać tylko jedną sztuczkę. Nawet jeśli sztuczka ta „ogólnieje” i z gry w szachy staje się umiejętnością uczenia się dowolnej gry (co jest *naprawdę* wspaniałym osiągnięciem!), nasz „program grający” nie tylko nie poprowadzi z nami konwersacji, ale też nie zorientuje się, że wokół szaleje pożar i wciąż będzie uprzejmie zachęcał do kolejnej rozgrywki. Choć można oczywiście napisać program wykrywający pożary i bota konwersacyjnego (choć te ostatnie wciąż są niestety żenująco kiepskie), to po ich dodaniu do zasobnika programów „mojego robota”, uzyskamy tylko trójkę sprzęgniętych programów: {SUPERGRACZ, POŻARATOR, GADULEC}.

Powyzłościwiałem się w poprzedniej sekcji na „szkołę dramatyczną”, jednak trudno opędzić się od myśli, że nawet bardzo długa lista programów wciąż będzie tylko listą programów, i od tęsknoty za owym tajemniczym „czymś więcej”. Powiem wam szczerze – osobiście uważam, że odpowiednio długa lista programów mogłaby stanowić *naprawdę znakomitą* imitację żywej, myślącej, przytomnej osoby. Mimo wszystko byłoby jednak fajnie, gdyby nasz robot otrzymał coś w stylu „minimalnej osobowości”, jakąś prostą protezę jaźni i dopiero na jej usługach byłyby rozmaite programy szczegółowe.

Jedną z prób stworzenia czegoś właśnie takiego są **architektury poznawcze/kognitywne** (*cognitive architectures*). To stary pomysł – można wręcz spotkać się z opinią, że „archaiczny” – jednak trudno się w nim nie zakochać.





Architektury kognitywne

Pozwólcie, że otworzę dłuższym cytatem z fenomenalnego artykułu przeglądowego^[26] na temat architektur^[14*] kognitywnych z 2020 roku. Mógłbym pewnie sparafrazować te słowa, ale to by było tylko udawanie, że wymyśliłem to wszystko sam. A więc:

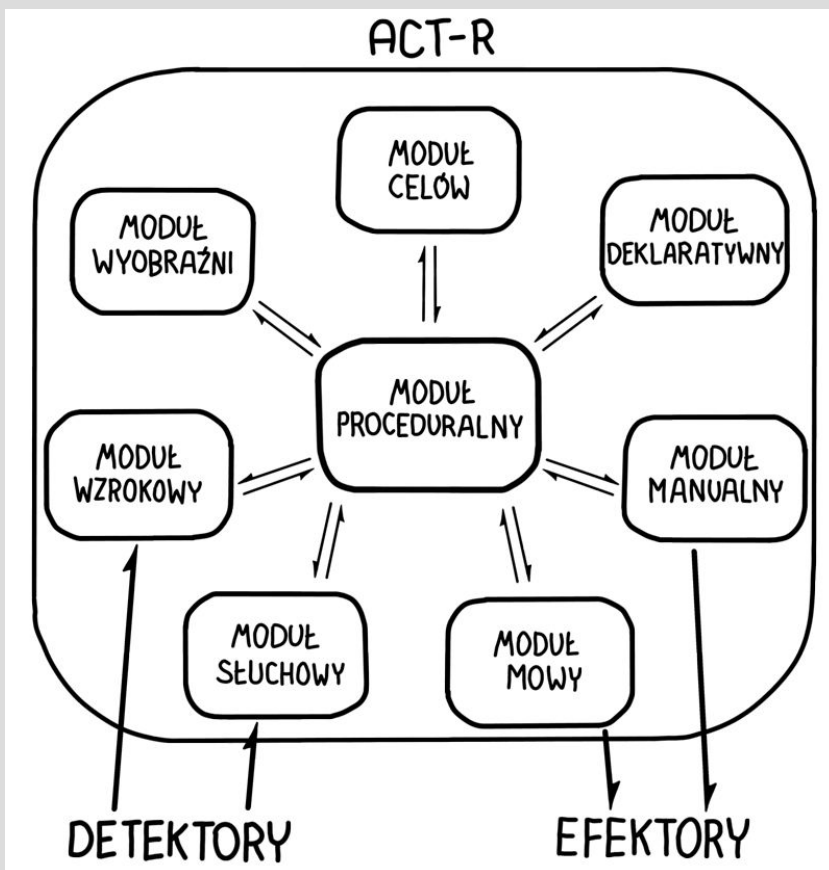
Architektury kognitywne są częścią badań nad ogólną sztuczną inteligencją, które rozpoczęły się w latach 50. XX wieku z myślą o stworzeniu programów mogących rozumować na temat problemów należących do różnych dziedzin, uzyskiwać wgląd w te problemy, adaptować się do nowych sytuacji i dokonywać refleksji nad sobą samym. Ostatecznym celem badań nad architekturami kognitywnymi jest modelowanie ludzkiego umysłu, co ma nam umożliwić skonstruowanie sztucznej inteligencji na poziomie ludzkim. W tym celu bada się, jakie konkretnie mechanizmy odpowiadają za inteligentne zachowanie [...] Choć nie istnieje pełna lista takich umiejętności, zwykle identyfikuje się szerokie obszary, takie jak [...]: percepcja, pamięć, uwaga, aktywność cielesna, oddziaływanie z ludźmi, planowanie, motywacja, emocje.

Choć istnieją „tematyczne architektury kognitywne” (mające na przykład symulować doświadczenie i tworzenie muzyki (MusiCog), rozwiązywanie problemów (FORR) albo wręcz wzrokową inspekcję powierzchni (ARDIS)), nas będą tu interesowały rozwiązania „ogólne”, czyli mające stanowić, nie bójmy się tego powiedzieć, zupełny model ludzkiego umysłu. Ważną cechą architektur kognitywnych jest to, że mają one z założenia funkcjonować w czasie rzeczywistym: nie określilibyśmy tym terminem programu komputerowego, który jest „uruchamiany na noc” i w tym czasie „myśli” sobie nad jakimś problemem, choćby nawet robił to „po ludzku”. Ma to być urządzenie, zwykle „ucieleśnione” – tj. podpięte pod różnego rodzaju czujniki i dysponujące różnymi „efektorami” typu koła, nogi czy ręce – spontanicznie reagujące na bieżąco na to, co mu się przydarza, na sposób jak najbardziej ludzki, czasem wręcz odruchowy (o czym niżej). Stopień „uczłowieczenia”, czy wręcz „uzwierżenia” architektury kognitywnej pozostaje kwestią fantazji twórców: nie jest niezwykle to, że modeluje się w nich zjawiska takie jak głód, pragnienie, niepewność, zaufanie czy strach.

Autorzy przeglądu, który cytowałem powyżej, szacują ostrożnie, że istnieje dziś około 100 aktywnych projektów tego typu, niektóre z nich rozwijane bezustannie od lat 70. Pośród najbardziej znanych, mających też najwięcej realnych zastosowań, można wymienić: Soar, ACT-R, EPIC, LIDA, CLARIOS, ICARUS, DIARC czy Disciple (podają te nazwy, bo może kiedyś pojawią się na ulotce waszego domowego robota pomocnika).

Istnieje oszałamiająca różnorodność architektur kognitywnych, a ja nie chcę męczyć was teorią. Dość powiedzieć, że niemal wszystkie z nich są w większym czy mniejszym stopniu inspirowane architekturą ludzkiego mózgu. Choć niektórzy mają nadzieję, że kiedyś uda się po prostu stworzyć jedną wielką sieć neuronową – odpowiadającą „sieciami neuronowymi” naszego mózgu – w praktyce jest to obecnie (i jeszcze przez długi czas pozostanie) z wielu względów niewykonalne. Czymś w rodzaju „modelu standardowego” stało się więc zaprojektowanie głównych modułów takiej architektury, a następnie realizowanie poszczególnych funkcji w mniej czy bardziej inteligentny sposób, często przy użyciu sieci neuronowych. O architekturach kognitywnych można więc pomyśleć jako o sposobie na "spięcie ze sobą" wszystkich najfajniejszych osiągnięć współczesnych badań nad AI, jak rozpoznawanie głosu i obrazu, generowanie mowy albo elastyczne sterowanie ruchem ciała.

Bywa, że inspiracja anatomiczna jest naprawdę silna. Oto zasadnicza struktura ACT-R, architektury kognitywnej rozwijanej nieprzerwanie od drugiej połowy lat 70. (!):



Mapa ta odpowiada w dużym uproszczeniu budowie mózgu ludzkiego. Ba, poszczególnym modułom przypisano nawet odpowiedniki anatomiczne. Przykładowo, twórcy ACT-R identyfikują swój moduł proceduralny z tzw. jądrami

podstawnymi, a moduł celów z tylnym zakretem obręczy. Tego typu proste przyporządkowania wywołują, nawiasem mówiąc, protesty neurologów, którzy znają prawdziwą złożoność mózgu i wiedzą, że nie składa się on z ostro oddzielonych od siebie modułów o prostych funkcjach. Prawdziwe piękno architektury kognitywnej – i drzemiąca w nich potęga – ujawnia się jednak dopiero wtedy, kiedy zrozumiemy, jak działają.

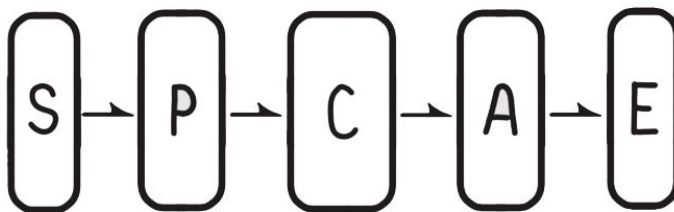


DIARC

Przyjrzymy się teraz DIARC: architekturze kognitywnej dość intensywnie rozwijanej od początku pierwszej dekady XXI wieku. Architektura ta ma jedną szczególną zaletę z naszego punktu widzenia: przeprowadzono eksperymenty z „zaszczepieniem” jej robotowi Nao, którego poznaliśmy już w roli „robotycznego trenera fitness” (zob. ramka na s. 18). Dzięki temu możemy sobie względnie łatwo wyobrazić „DIARC w działaniu”.

DIARC jest dynamicznie rozwijanym projektem i doczekał się wielu konkretnych „wcieleni”. Ja posłużę się tutaj tylko jedną z architektur DIARC, opisaną w 2006 roku [\[27\]](#).

Zacznijmy od podstaw. Oto ogólna „mapa” DIARC; strzałki pokazują kierunek przepływu informacji.

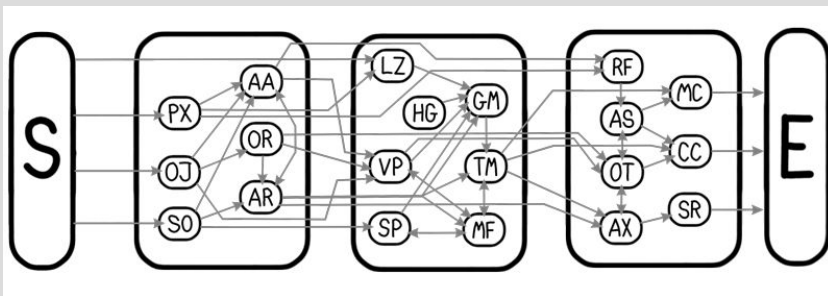


- S (*sensors*): sensory/czujniki – to wszystkie źródła „bodźców” dla DIARC, na przykład czujnik poziomu baterii, sonar, laser, kamera, mikrofon.
- P (*perceptual processing*): moduł percepcyjny – tutaj wykonywana jest pierwsza tura analizy „zmysłowej”: wykrywanie obiektów i odległości od nich, identyfikacja dźwięków i tak dalej.
- C (*central processing*): moduł centralny – tutaj na podstawie sygnału ze zmysłów oraz pamięci, modułu emocji, a także głębokich i tymczasowych celów sformułowany zostaje „plan zadaniowy”.
- A (*action processing*): moduł planowania działań – tutaj zadanie, które w module C wyznaczono do wykonania, zostaje „przetłumaczone” na konkretne działanie, takie jak ruch albo produkcja mowy.

- E (effectors): efekторы – to sterowniki kół, rąk, nóg, kamer, głośnika generującego mowę i tak dalej.

Wszystko to brzmi, mam nadzieję, bardzo sensownie. Aż nazbyt sensownie. Można by wręcz pomyśleć, że trudno wyobrazić sobie robota funkcjonującego *inaczej*. Żeby zrozumieć, jak subtelne jest w istocie działanie DIARC, musimy zejść o jeden poziom głębiej. Nie przejmujcie się, jeżeli poniższy diagram, nawet po moich wyjaśnieniach, wciąż pozostanie nie do końca jasny. Tak naprawdę, aby w pełni docenić rolę poszczególnych modułów i ich zależności, należałoby zejść jeszcze nieco głębiej i przyjrzeć się sposobom, w jaki one działają i jakie przekazują sobie dane. Celem tej ramki jest zaś głównie inspiracja.

Wrzucam więc pełen diagram i parę komentarzy do niego, ale bez ambicji, że daję wam do rąk faktyczny „przewodnik po DIARC” (którego zresztą nie umiałbym napisać, bo każda kolejna godzina spędzona przy dokumentacji tej architektury prowadzi mnie tylko głębiej w arkania inżynierii, bez nadziei na szybkie oświecenie). Nic to, jedziemy!



AA – Affect Appraisal (Ocena afektu)
 AR – Affect Recognition (Rozpoznawanie afektu)
 AS – Action Selection (Wybór działania)
 AX – Affect Expression (Wyrażanie afektu)
 CC – Camera Control (Kontrola kamery)
 GM – Goal Manager (Menedżer celów)
 HG – High-Level Goals (Cele wysokiego poziomu)
 LZ – Localization (Lokalizacja)
 MC – Motion Control (Kontrola ruchu)
 MF – Memory (Pamięć)

OJ – Object Detection (Wykrywanie obiektów)
 OR – Object Recognition (Rozpoznawanie obiektów)
 OT – Object Tracking (Śledzenie obiektów)
 PX – Proximity (Bliskość)
 RF – Reflexes (Odruchy)
 SO – Sound Detection (Wykrywanie dźwięku)
 SP – Speech Processing (Przetwarzanie mowy)
 SR – Speech Production (Generowanie mowy)
 TM – Task Manager (Menedżer zadań)
 VP – Visual Processing (Przetwarzanie obrazu)

Od czego by tu zacząć... Idźmy może od lewej strony.

Trzy „pierwsze” moduły po lewej stronie stanowią pierwszą warstwę „interpretacyjną” dla surowych sygnałów z czujników, tłumacząc je na bardziej użyteczną informację o tym, jakie obiekty znajdują się w środowisku (OJ), jaka jest odległość od nich (PX) oraz jakiego typu dźwięki występują w środowisku (SO). Nie będę już tego odrębnie pisał o każdym module z osobna, ale nawet tego typu „proste” zadania potrafią być szalenie wymagające. Ot, wydzielenie w ciągłym sygnale z kamery poszczególnych obiektów (OJ) i ich identyfikacja (OR) – to tak naprawdę skomplikowane, nietrywialne zadanie, które prawdopodobnie realizować będzie złożona sieć neuronowa rozwijana latami przez wyspecjalizowaną grupę inżynierów. Zostawmy to i wędrujemy dalej.

Dość szybko natrafiamy na pierwszą ciekawostkę. Zwróćcie uwagę na strzałki prowadzącą od modułów PX, OJ i SO do modułu AA (*Affect Appraisal*, czyli Ocena afektu). W jego nazwie pojawia się słowo „afekt” – termin psychologiczny, który profani tacy jak ja tłumaczą sobie szybko w głowie na „emocja” albo „nastój”. Przykładowo, „zaburzenia afektywne” określa się też czasem jako „zaburzenia nastroju”, a należą do nich m.in. depresje, manie czy zaburzenia dwubiegunowe. DIARC posiada więc odpowiednik „stanu emocjonalnego”. Co z tego wynika? Jak można z niego skorzystać?

Podążajmy za strzałkami. Jedna z nich prowadzi, w towarzystwie strzałki od modułu PX, do modułu RF („odruchy”), omijając przy tym cały blok centralny (w którym znajdują się moduły odpowiadające choćby za pamięć, planowanie czy wyższe cele). Oznacza to w praktyce, że sygnał dotyczący bliskości jakiegoś przedmiotu albo negatywny ładunek emocjonalny mogą wywołać działanie z pominięciem „wyższych” funkcji poznawczych. Dzięki tym dwóm strzałkom możliwe jest więc choćby zakodowanie odruchowego, szybkiego uchylenia się przed rzuconą piłką („łuk” PX-AA-RF) – nie ma czasu sięgać do pamięci i analizować hierarchii celów, gdy trzeba ratować skórę.

„Afekt” wykrywany jest też w sygnale dźwiękowym, na przykład w czyimś głosie (za to odpowiada z kolei moduł AR). Emocje kryjące się w głosie ludzkim również mogą więc wywoływać reakcję z pominięciem bloku centralnego i tylko od fantazji programistów zależy, czy na listę trafi wyłącznie groźny, głośny wrzask (wywołujący w module AS, czyli „selektorze działań”, na przykład odruchowe odskoczenie), czy też na listę trafiają ponadto dźwięki lub słowa wywołujące – równie automatycznie i poza kontrolą „bloku centralnego” – rozbawienie, zażenowanie albo wstyd.

Spójrzmy teraz na niewinny moduł AX w bloku A, który „czerpie inspirację” z innych modułów afektywnych, a wpływa z kolei na wybór działania (AS) i emisję głosu (SR). Dzięki niemu aktualny stan „emocjonalny” robota („jestem spokojny”) albo „klimat emocjonalny środowiska” („znajduję się w groźnym miejscu”) mogą decydować o podejmowanych przez niego czynnościach i wpływać na ton głosu. To w tym „łuku” można więc zaprogramować wiedzę choćby o tym, że gry i zabawy mają „sens” tylko w określonym stanie emocjonalnym (a więc na pewno nie wtedy, kiedy płonie dom, ale raczej również nie wtedy, kiedy ktoś w tym samym pokoju płacze) albo że w obecności skaczącego jak piłka przedszkolaka mówi się szybciej i radośniej niż w obecności zadumanego starca. Co istotne, zasady te mogą mieć charakter bardzo ogólny, na przykład w oparciu o stosowaną przez psychologów ideę „walencji”, dzielącą wszystkie emocje na osi „pozytywne–negatywne”. Dzięki temu robota nie trzeba byłoby żmudnie uczyć wszystkich możliwych kombinacji emocji i czynności, a raczej ogólnej zasady „dopasowywania się do nastroju”.

Jak widać, w tym względnie prostym (choć pewnie i tak wyglądającym na pierwszy rzut oka jak schemat elektrowni jądrowej) schemacie kryje się naprawdę wiele subtelności. Odkrycie niektórych z nich pozostawiam wam – naprawdę gorąco zachęcam do „przespacerowania się” po tym schemacie i zastanowienia, jakie funkcje umysłu można by zrealizować przy pomocy danych strzałek. Obiecuję, że podziękujecie mi, kiedy już dostaniecie swojego pierwszego robota domowego wyposażonego w zaawansowaną architekturę kognitywną. Z mojej strony tylko ostatni przykład.

Wielką bolączką twórców robotów jest umiejętność zaprogramowania w nich zdolności do samodzielnego inicjowania działań i decydowania, co jest w danej sytuacji najważniejsze. Anegdotyczną, ale jakże klarowną demonstracją tego problemu jest komputer, który postanawia 5 minut przed ważną prezentacją dokonać pełnej aktualizacji systemu, co zajmuje 30 minut. Jest to jednak problem ogólniejszy:

AI wydaje się pilnie potrzebować „zdrowego rozsądku”, pozwalającego na wybieranie działań, kierując się jakimiś ogólnymi zasadami.

Zobaczmy może, jakie czynniki wpływają na to, czym zajmie się w najbliższym czasie robot sterowany przez DIARC, a więc – mówiąc precyzyjniej – jaki cel (*goal*) wybierze do realizacji moduł GM (Menedżer celów). Dodajmy, że dla DIARC słowo „cel” to względnie wysoki poziom postawienia problemu. Twórcy tej architektury określają go jako „docelowy stan środowiska zewnętrznego lub wewnętrznego, który należy uzyskać”. Przykładowe cele to: „uzupełnić kieliszek Marysi” albo „posprzątać pokój”, choć zauważmy, że cele mogą też być „wewnętrzne”, na przykład „dowiedzieć się, gdzie w Krakowie znajduje się ulica Wiślna” albo „skomponować sprośny limeryk na temat pewnego mężczyzny z miasta Löbau”.

Niższy poziom organizacji to „zadanie” (*task*). To już konkretna czynność, na przykład „podejść do lodówki”, „otworzyć drzwi”, „wyciągnąć butelkę soku pomarańczowego”. Moduł GM tłumaczy więc wybrany do realizacji cel („uzupełnić kieliszek Marysi”) na konkretne zadania, zaś moduł TM (Menedżer zadań) tłumaczy konkretne zadanie („podejść do lodówki”) na konkretne polecenia dla sterowników: „obróć się nieco w prawo i ruszaj naprzód”.

Zobaczmy teraz, skąd DIARC czerpie „inspiracje” przy wyborze celu do realizacji. Jedna strzałka do modułu GM pochodzi od modułu SP (Przetwarzanie mowy). Tędy spływają więc polecenia głosowe: „posprzątaj pokój” albo „otwórz drzwi kapsuły, HAL”.

Inna prowadzi od modułu AR, który odpowiada za wykrywanie „emocji innych”. Istnienie tej ścieżki pozwala reagować na „atmosferę” w pokoju. Przykładowo, DIARC może uznać, że ze względu na ogólny poziom rozradowania osób w pomieszczeniu spośród dwóch potencjalnych celów – posprzątania pokoju i uzupełnienia kieliszków – warto zacząć od tego drugiego.

Spójrzmy jednak na strzałkę prowadzącą od modułu HG (Cele wysokiego poziomu). Twórcy DIARC są zaskakująco skromni w wyjaśnianiu, czym są owe „wyższe cele”. Wygląda na to, że moduł ten można wyposażać we wszystko, co nam tylko fantazja podpowie, od stałej listy elementarnych celów egzystencjalnych („nie ulec zniszczeniu”, „nie rozładować baterii”), przez cele zahaczające o filozofię robotyki („nie krzywdzić ludzi”, „nie sprzeciwiać się właścicielowi”), aż po listę czysto zadaniową („utrzymywać dom w czystości”). Tak czy inaczej, są to cele „stałe obecne w tle” – Menedżer celów zawsze o nich „wie”, dzięki czemu może stale dbać o to, aby cel spływający z modułu SP („uzupełnij kieliszek Marysi”) nie był w sprzeczności z którymś permanentnym celem z modułu HG („nie krzywdzić ludzi”). Gdybyście się zastanawiali – tak, istnieją już badania nad programowaniem w robotach mechanizmów sprzeciwu i tego, jak go najskuteczniej wyrazić (głos jest skuteczniejszy od tekstu, prośba jest skuteczniejsza od rozkazu).

Na tym przerwę – nie dlatego, że wyczerpaliśmy już wszystkie subtelnosci DIARC, a raczej dlatego, że nie chcę wyczerpać was.

Co z tego wszystkiego wynika? Zauważmy przede wszystkim, że jest to tylko „architektura”, o której można pomyśleć jako o „miejscu na moduły” – i to dopiero od jakości wykonania tych modułów będzie zależeć poziom „inteligencji” danego robota. Na cóż nam nawet najpiękniejsza architektura kognitywna, jeśli moduł OJ nie będzie w stanie zauważyć na stole kieliszka, moduł GM uzna, że najlepszym

sposobem posprzątania pokoju będzie wyrzucenie wszystkich książek do śmieci, a moduł MC każe naszemu robotowi wiecznie potykać się o dywan? Wnioski? Zawsze – zawsze! – diabeł tkwi w szczegółach. Nie spodziewajcie się, że „sztuczna inteligencja ogólna” sfrunie z nieba w postaci gotowego, działającego rozwiązania. Gdzieś tam zawsze będzie siedziała ekipa sfrustrowanych inżynierów, którzy od dwóch miesięcy trują się, aby ich robot prawidłowo odróżniał psa od kota.

Przypuśćmy jednak, że zechcemy odłożyć te wszystkie „brudne” problemiki na bok i założymy niefrasobliwie, że po prostu kiedyś zostaną rozwiązane. Co wtedy? No właśnie. Początkowo chciałem pokazać działanie architektur kognitywnych na jakimś odpowiednio spektakularnym przykładzie – robota, który ujawniałby głębię swoich procesów myślowych. Okazuje się jednak, że twórcy robotów z architekturami kognitywnymi na co dzień zmagają się głównie z owymi „drobnymi problemikami” i chwalą się raczej postępami w ich rozwiązywaniu niż rzeczywistym funkcjonowaniem robotów posiadających „umysł”. Architektury kognitywne rozwijają się więc raczej w tle rzeczywistego postępu w robotyce i cieszą względnie małym zainteresowaniem. Póki co poziom inteligencji robotów nie wydaje się ograniczony złożonością struktury ich „umysłów”, a raczej stopniem wyrafinowania poszczególnych modułów.

Co więcej, ową głębię umysłu niełatwo zademonstrować. Wiele bardzo inspirujących wyników może na pierwszy rzut oka prezentować się błado. Twórcy architektur kognitywnych mogą być bardzo dumni z tego, że sterowany nimi robot samodzielnie wybrał cel do realizacji i „zadecydował” o posprzątaniu stołu albo odruchowo odsunął się od płomienia, ale dla postronnego obserwatora wciąż jest to tylko robotyczne ramię machające zmiotką.

Ponieważ jednak nie chcę was zostawiać z pustymi rękami, przyjrzyjmy się na koniec dwóm ciekawym zastosowaniom architektur kognitywnych.



Dempster i Shafer^[28]

Dempster i Shafer to dwa roboty (nazwane na cześć dwóch matematyków, którzy badali rozumowanie w warunkach niepewności) wyposażone w „umysł” oparty na architekturze kognitywnej DIARC oraz „ciało” robota Nao. W 2017 roku opisano ciekawy eksperyment z ich udziałem – ciekawy przynajmniej z dwóch powodów.

Po pierwsze, była to para robotów ze wspólnym „umysłem”. Wszystkie informacje dostępne jednemu były też automatycznie dostępne dla drugiego. Tak naprawdę, jeśli się nad tym zastanowić, był to więc po prostu „jeden robot w dwóch częściach”. Z technologicznego punktu widzenia jest to trywialny krok, jednak jego konsekwencje są naprawdę interesujące. Wątek „zlewania się wielu maszyn w jedno” będzie przewijał się w tej książce (zwłaszcza w rozdziale III), warto więc pamiętać przy tych okazjach, że granica pomiędzy „wieloma dobrze skomunikowanymi maszynami” a „jedną maszyną składającą się z rozłącznych fizycznie elementów” jest czysto językowa – a jednak jakże znacząca. Zupełnie inaczej myśli się o „miliardzie komputerów rozsianych po świecie” i „jednym wielkim globalnym komputerze”.

Po drugie, eksperyment ten pozwolił zademonstrować kilka głębokich cech DIARC. Niestety, niewiele tych cech przejawiało się na poziomie widocznych zachowań i interesujące jest raczej to, co dzieje się „pod maską”. Inaczej mówiąc, dużo prostsze roboty mogłyby zachowywać się dokładnie w ten sam sposób. Spróbujmy jednak wczuć się w klimat i docenić, że w tym konkretnym przypadku owe proste zachowania wynikły jednak z naprawdę ciekawych reguł.

Przykładowo, robot poproszony o to, by iść przed siebie, robił to tylko wtedy, jeśli „czuł się pewnie”, odmawiając wykonania polecenia, gdy znalazł się zbyt blisko krawędzi stołu. Choć na pozór mogłoby się wydawać, że jest to zwykły przypadek wykrywania krawędzi, w istocie kryje się za nim bardziej złożona „psychologia”. Robot odmawiał wykonania polecenia nie dlatego, że miał sztywno zakodowane zatrzymywanie się w obliczu krawędzi (co potrafiłyby zrobić byle prosty robocik z Lego), a raczej dlatego, że pojawił się konflikt pomiędzy celami: polecenie głosowe operatora kłóciło się z „głębokim” celem „nie ulec uszkodzeniu”. Innymi słowy, za decyzją, aby odmówić wykonania polecenia operatora, kryły się „wartości” (zakodowane w głębokich celach, czyli – odsyłam do powyższej ramki – module HG) oraz „emocje” (zakodowane w module AA).

Co więcej, konflikt ten można było rozwiązać: operator jest w stanie na przykład poinformować robota, że ma iść dalej, ponieważ w razie upadku ze stołu zostanie złapany. Robotowi zakodowano ideę łapania i potrafił odnieść ją do spadania ze stołu („złapanie przez operatora przerywa spadanie”), dzięki czemu ostatecznie przypisywał czynności „pójścia naprzód” znacznie bardziej pozytywny afekt. Mówiąc nieformalnie, zaufanie do operatora pozwoliło przezwyciężyć strach. A mówiąc precyzyjniej, wykonanie zadania „iść naprzód” nie stało już w sprzeczności z celem „nie ulec zniszczeniu”, ponieważ moduł pamięci podpowiadał, że w tym przypadku nie dojdzie do zderzenia z podłogą. Ostatecznie robot posłusznie ruszył, „ufając”, że zostanie złapany.

Co ciekawe, w niektórych architekturach kognitywnych obecny jest samouczący się model pamięci, oparty na sieciach neuronowych, który zupełnie spontanicznie zakodowałby wynik takiego eksperymentu jako „pozytywnie się kojarzący” – gdyby robot rzeczywiście został złapany – lub „negatywnie się kojarzący” – gdyby

operator postanowił zrobić robotowi głupi dowcip. Mówiąc metaforycznie, robot taki musiałby się dopiero „nauczyć zaufania”.

Ważne jest, by zrozumieć, że w sieciach neuronowych nie dokonuje się to poprzez wpisanie jakiejś wartości w określoną komórkę „bazy danych zaufania”, odpowiadającą na przykład „poziomowi zaufania do operatora” – co byłoby rozwiązaniem archaicznym i szalenie mało elastycznym (już choćby z tego względu, że ktoś musiałby wcześniej przewidzieć istnienie tej konkretnej zmiennej). Mówimy tu o trenowaniu sieci neuronowych metodą uczenia maszynowego (zob. ramkę na s. 24), czyli zjawisku podobnym do rzeczywistego biologicznego uczenia się. Sieć neuronowa sprzed „fatalnego wypadku” i po jego nastąpieniu różni się, ponieważ zdarzenie to wywarło wpływ na stan jej neuronów. Jest to jednak wpływ „rozproszony” pomiędzy tysiące neuronów, które w rezultacie nieco inaczej niż wcześniej przetwarzają sygnał (choć ostatecznie, rzecz jasna, również mówimy o zmianie wartości pewnych zmiennych). Różnica między klasycznym podejściem „eksperckim” a uczeniem maszynowym jest jednak subtelna i będzie wielokrotnie przewijała się w tej książce, prawdopodobnie nigdy nie doczekawszy się jednoznacznego wyjaśnienia. Na chwilę obecną wystarczy, abyśmy zrozumieli, że uczenie maszynowe prowadzi naturalnie do pewnych ciekawych zjawisk poznawczych, które znamy doskonale z pierwszej ręki jako posiadacze mózgów.

Prosty przykład: robot, który wykryłby którymś ze swoich czujników, że wskutek upadku doszło do uszkodzenia jego „ciała”, a ponadto miał zakodowaną w głębokich celach informację, że uszkodzenie siebie jest czymś negatywnym, w przyszłości reagowałby negatywnym „afektem” po prostu na wszelkie odpowiednio podobne sytuacje – i zupełnie jak strauumatyzowany człowiek mógłby na przykład „zapamiętać” przy okazji kolor koszulki albo grającą w tle muzykę i reagować na nie „irracjonalnym niepokojem”.

To nie jest tylko teoria.



O robocie, który się bał^[29]

W tym samym 2017 roku opisano prostą architekturę kognitywną mającą symulować „uczenie się strachu”. Wyposażono ją w moduł analizy obrazu i dźwięku, pamięć roboczą oraz odpowiedniki niektórych ludzkich struktur neuronalnych odpowiedzialnych za warunkowanie i lęk. Autorzy, silnie zainspirowani ludzką neurologią, posługiwali się terminami takimi jak „moduł hipokampu” i „moduł ciała migdałowatego”, zaś poziom lęku reprezentowany był przez zmienną o nazwie „adrenalina”. Architekturę tę „zaszczepiono” następnie robotowi – znów był to Nao! – po czym rozpoczęto okrutne eksperymenty.

Przed wszystkim należało ustalić, co będzie „naturalnie nieprzyjemne” dla robota. Postawiono na ciemność: im ciemniejszy był obraz rejestrowany przez Nao, tym wyższy był poziom lęku. Następnie wykonano serię prostych eksperymentów odpowiadających klasycznemu warunkowaniu. Prezentowano bodziec „neutralny”, na przykład ludzką twarz, logo NAO albo różne dźwięki, a niedługo po nim ciemność, wywołującą skok adrenaliny. Jak łatwo się domyślić, po pewnym czasie sieci neuronowe Nao reagowały „niepokojem” również na bodźce neutralne, jeśli tylko odpowiednio przypominały te, którym wcześniej towarzyszyła ciemność. Wyhodowano więc robota, który reagował negatywnie na dźwięki „nieprzyjemnie się kojarzące”. Potrafił się też spontanicznie nauczyć, że jeśli jakiś bodziec czasem występuje wraz z ciemnością, a czasem nie, nie należy się go „bać”, ponieważ powiązanie to jest najwyraźniej przypadkowe.

Łatwo sobie wyobrazić, że „wszczepienie” tego typu modułu do, powiedzmy, architektury DIARC mogłoby sprawić, że robot poproszony o przyniesienie słoika z ciemnej piwnicy początkowo odmówi, a następnie – jeśli zostanie odpowiednio zachęcony i przemoże swój lęk – może wyrobić sobie dodatkowy odruch lękowy w reakcji na coś, co ujrzał w tej piwnicy, na przykład odkurzacz. Inny przykład: robot domowy posiadający tego typu architekturę oraz moduł rozpoznawania twarzy mógłby się spontanicznie nauczyć negatywnych bądź pozytywnych skojarzeń z konkretną osobą. Ponieważ „moduł afektu” wpływa pośrednio na wszystkie dalsze czynności, robot taki mógłby różnie się zachowywać w obecności różnych osób.

I.2.2. Uczenie się

Wróćmy teraz do drugiego problemu, który zapowiedziałem wcześniej: elastycznego uczenia się.

W rozdziale tym – i w ogóle w całej tej książce – przewijają się dwa modele tworzenia „sprytnych maszyn”, czasem przedstawiane przeze mnie jako sobie przeciwne. Jednym jest **system ekspercki**: program

zostaje z góry zaprogramowany i wyposażony w sformułowane przez ludzkich ekspertów „cegiełki wiedzy”. Drugim jest **uczenie maszynowe**: program komputerowy rozpoczyna jako *tabula rasa* – „czysta kartka” – i samodzielnie nabiera kompetencji w toku tysięcy i tysięcy prób.

Problem z pierwszą metodą jest taki, że wymaga olbrzymiego nakładu pracy ekspertów, a przypadki nieprzewidziane przez nich potrafią „przelecieć przez sito”. Problem z drugą metodą jest taki, że wymaga podejmowania olbrzymiej liczby prób. Pozwólcie, że to podkreślę: *naprawdę* olbrzymiej. AlphaZero, wiodący obecnie algorytm szachowy, osiągnął niebywały wręcz poziom kompetencji szachowych, startując „od zera”, to jest tylko od krótkiej listy elementarnych reguł gry. Następnie grał uparcie z różnymi wersjami „siebie”, zawsze starając się wygrać. Do osiągnięcia poziomu mistrzowskiego potrzebował, bagatela, 44 milionów partii.

Pomyślmy chwilę o tej liczbie. Nie stanowi ona większego problemu, gdy mówimy o komputerze pozostawionym samemu sobie, jednak gdy przedmiotem treningu ma być relacja człowiek–maszyna, biedna maszyna musiałaby co chwilę czekać na powolną jak lodowiec reakcję swojego biologicznego partnera. W przypadku szachów – pobawmy się może trochę liczbami – oznaczałoby to, że armia 100 szachistów grająca bez wytchnienia, dzień i noc, szybciućkie partie trwające średnio łącznie 5 minut potrzebowałaby ponad czterech lat na ukończenie takiego treningu. (A i tak w niedługim czasie okazałoby się, że nawet najlepsi szachiści świata przestają stanowić wyzwanie dla algorytmu).

W praktyce konieczne jest więc łączenie tych dwóch technik. Przykładowo, roboty szkoli się wstępnie w wirtualnej rzeczywistości, co pozwala na wykonywanie tysięcy i tysięcy prób bez ograniczeń czasowych i zniszczeń sprzętu, a następnie „ostatnie szlify” przeprowadza się już w świecie realnym (zob. ramka na s. 28). Przypuśćmy więc – jak fantazjowaliśmy pod koniec rozdziału I.1 – że robot domowy trafia do naszego mieszkania wstępnie wyszkolony, jednak konieczne okazuje się „dotarcie”. Wygląda na to, że na tym etapie potrzebne jest coś nieco „szybszego” niż klasyczne sieci neuronowe. Nie chcemy robota, który sto razy sięgnie po zły kubek, zanim w końcu nauczy się na błędach (na przykład obserwując naszą zirytowaną twarz), co właściwie mamy na myśli, mówiąc „ten z obtłuczonym uchem”.

W całym oceanie możliwych rozwiązań tego problemu opowiem o dwóch tropach: samodzielnym dopytywaniu przez robota i uczeniu się przez demonstrację.



Robot ciekawski^[30]

W 2017 roku opisano robota, którego zadaniem było podawanie użytkownikowi jednego spośród wielu dostępnych mu przedmiotów. Haczyk polegał na tym, że robot ten startował z zerową wiedzą, posiadał natomiast zdolność zadawania pytań – testowano, jaka metoda zdobywania wiedzy będzie najbardziej skuteczna i jaka najbardziej przyjazna dla człowieka.

Użytkownik mógł więc, przykładowo, rozpocząć od komendy typu „Podaj mi niski słoik z niebieską zakrętką”. Co ważne, założono „słownik otwarty”, tj. nie ograniczono słownictwa robota do zamkniętej listy „użytecznych” przymiotników lub rzeczowników. Dzięki temu, gdyby użytkownik konsekwentnie stosował dowolne rzeczywiste słowo^[15*] – nawet rzadkie lub niejednoznaczne jak „tandetny”, „połykliwy” albo „peruwiański” – to robot włączyłby je do swojego słownika.

Robot miał do dyspozycji dwa sposoby poproszenia użytkownika o wyjaśnienie:

- „Czy mógłbyś wskazać ręką na obiekt, który określiłbyś słowem *p*?” [robot śledzi kamerą ruchy człowieka]
- „Czy użyłbyś słowa *p* do opisania *tego* przedmiotu?” [robot wskazuje swoją „ręką” na wybrany przedmiot i czeka na odpowiedź]

Po osiągnięciu zadowalającego poziomu pewności robot wskazywał na jeden z przedmiotów i pytał: „Czy miałeś na myśli *ten* przedmiot?”.

A więc mamy tutaj do czynienia z typowym przypadkiem uczenia poprzez przykłady. Nie jest to wcale proste zadanie, przynajmniej dla człowieka – wyobraźmy sobie, że znajdujemy się na miejscu robota, a ktoś prosi nas o obiekt, który jest jednocześnie „bulu”, „kekere” i „wuwo”, a za wyjaśnienie tych pojęć musi nam wystarczyć wskazanie palcem na przykładowy przedmiot. Trzeba naprawdę niezłej pamięci i sprawnego operowania „mentalnymi tabelkami”, żeby uczyć się w ten sposób rozpoznawania większej liczby przedmiotów. Szczęśliwie, jest to akurat przykład zadania względnie prostego dla komputera. Jedyny problem polega na tym, jak zoptymalizować proces uczenia. Najprostszą metodą byłoby wszak zapytanie kolejno o każdy spośród 10 przedmiotów, czy jest on „bulu”, i tak dalej. Byłoby to jednak szalenie nużące dla użytkownika. Problemem do rozwiązania jest więc stworzenie „stylu uzyskiwania wiedzy”, który jest nie tylko skuteczny, ale i przyjazny człowiekowi.

Autorzy porównali dwa algorytmy: „zorientowany na zadanie” i „ciekawski”.

Robot „zadaniowy” pytał zawsze tylko o to, co mu było potrzebne do wykonania bieżącego zadania – po „rozłożeniu” wypowiedzi użytkownika na czynniki pierwsze identyfikował w nich słowa klucze, po czym pytał kolejno o wszystkie, których nie umiał jeszcze odnieść do rzeczywistości, od czasu do czasu próbując odgadnąć właściwy przedmiot. W podanym wyżej przypadku mógłby więc, przykładowo, poprosić kolejno o 1–2 przykłady przedmiotów niskich, stoików, niebieskich i zakrętek, a potem wykonać pierwszy „strzał”. W toku kolejnych prób słownik szybko stawał się coraz bardziej „nasycony”: w prośbach użytkowników coraz rzadziej pojawiały się słowa nieznanne. Ze względu na ograniczoną liczbę przedmiotów coraz częściej udawało się też odgadnąć, który z nich użytkownik ma na myśli – nawet jeśli nie każde słowo zostało wyjaśnione.

Robot „ciekawski” miał do dyspozycji dodatkową sztuczkę: mógł na dowolnym etapie testu zadać pytanie o dowolne słowo, które się dotychczas pojawiło – nawet jeżeli nie było potrzebne do rozwiązania bieżącego problemu. Zwykle działało się to wtedy, gdy opis użytkownika był dla robota dość jasny: oznaczało to, że można „poświęcić chwilkę” na dopytanie o jakieś słowo, które pojawiło się już wcześniej. Jest to, nawiasem mówiąc, znana sztuczka. W internecie można zagrać sobie w „grę w 20 pytań” – w momencie pisania tej książki choćby pod adresem www.20q.net. Program zadaje pytania, na które można odpowiadać tylko „tak” i „nie”, po czym zgaduje, o czym pomyśleliśmy. Bywa, że z zaskakującą, wręcz niepokojącą skutecznością. Potęga tych algorytmów leży w długim treningu, ale również w tym, że gdy algorytm osiąga zadowalający poziom pewności w okolicach pytania nr 10, poświęca pozostałych 10 pytań na „dokszałcenie się”.

Wróćmy jednak do eksperymentu. Po „przepuszczeniu” robota przez 30 użytkowników okazało się, że ciekawość się opłaca – i to pod każdym względem. Robot „ciekawski” był skuteczniejszy: zgadywał poprawnie niemal dwukrotnie częściej. Wiedza uzyskana na etapie „swobodnego dopytywania” przydawała się później, co nie jest aż tak zaskakujące: skoro ktoś użył wcześniej jakiegoś słowa, jest spora szansa, że użyje go w przyszłości. Co jednak istotniejsze, robot ciekawski uznawany był też jako bardziej interesujący i „zabawny”, a ponadto użytkownicy dawali mu wyższe noty w kategoriach „Robot wydawał się rozumieć moje słowa” oraz „Skorzystałbym z takiego robota, aby przyniósł mi coś z drugiego pokoju”. Co jeszcze ciekawsze, *nie* był oceniany wyżej w kategorii „Robot zadawał za dużo pytań”: owe dodatkowe pytania nie stały się jeszcze obciążeniem.

Nie wszystko da się wyrazić słowami. Mało tego, gdy przychodzi do wyjaśniania, czym jest wiedza, filozofowie tradycyjnie wydzielają dwa jej fundamentalne typy: „wiedzę *że*” i „wiedzę *jak*”. Ta pierwsza obejmuje wszystko, co możemy wypowiedzieć na głos: „wiem, że Warszawa jest stolicą Polski” i „wiem, że cukier jest na górnej półce”. Tego typu wiedzę względnie łatwo przekazać drugiemu człowiekowi – albo komputerowi. Druga kategoria obejmuje przypadki, które w zwykłym języku podciągnęlibyśmy raczej pod „umiejętności” niż „wiedzę”, na przykład „umiem pływać” i „wiem, jak wymienić wycieraczkę w samochodzie”. Nietrudno zauważyć, że przekazanie komuś tego typu umiejętności słowami bywa albo niemożliwe, albo skrajnie trudne. Aby uzyskać tego typu wiedzę, często trzeba

popatrzeć na eksperta przy pracy albo, najlepiej, samodzielnie wykonać pewne czynności: ponieważ żadne słowa nie wyrażą tego, z jaką siłą i pod jakim kątem trzeba nacisnąć ten mały plastikowy zypcyk (zwany też wihajstrem) od wycieraczki, żeby wyskoczyła z uchwytu.

Jeżeli roboty rzeczywiście mają się stać naszymi domowymi pomocnikami, będą musiały pilnie przyglądać się naszym czynnościom i uczyć się na podstawie obserwacji. Już teraz przyzwyczajacie się więc, że będą zaglądać nam przez ramię i zadawać mnóstwo głupich pytań – jak dzieci. W robotyce mówi się zwykle o trzech typach uczenia przez demonstrację^[31]:

- uczenie kinestetyczne (operator porusza fizycznie robotem),
- uczenie przez zdalne sterowanie (operator steruje robotem zdalnie),
- uczenie przez bierną obserwację (robot obserwuje operatora).

Jak to zwykle w bajkach bywa, każda z metod ma swoje wady i zalety.

Uczenie kinestetyczne jest bardzo intuicyjne dla człowieka – wyobraźmy sobie, że stajemy za robotem i ruszamy jego „rękami”, pokazując mu, jak się ściera kurz, sieka szczypiopek albo wymienia baterie w pilocie do telewizora (mam przecucie, że nawet z robotem pomocnikiem, podłączonym na stałe do telewizora, pilot nie odejdzie tak szybko w przeszłość). Problem polega na tym, że „ręce” (czy inne efekторы) robotów z reguły reagują inaczej niż ciało ludzkie i trzeba sporo wprawy, aby wymusić na nich odpowiedni typ ruchu. Zwykle okazuje się też, że trajektorie i sekwencje ruchów uzyskane w trakcie demonstracji trzeba na różne sposoby „wygładzać” i uogólniać.

Uczenie przez zdalne sterowanie jest przede wszystkim bardzo „bezpośrednie”, tj. możemy wymusić na robocie dokładnie taki ruch, jaki sobie wymarzymy, na przykład przesunięcie małego palca o milimetr w bok. Technicznie, oznacza to, że mamy dostęp do każdego „stopnia swobody” robota. Wymaga to jednak oczywiście dodatkowych kompetencji: gdybym miał taką metodą uczyć mojego domowego pomocnika, potrzebowałbym dodatkowego sprzętu kontrolującego jego ruchy, dodatkowego rozdziału w instrukcji obsługi i pewnie sporo wprawy.

Uczenie przez bierną obserwację (albo „uczenie przez imitację”) to w pewnym sensie ideał – wykonuję po prostu daną czynność zupełnie spontanicznie, nie musząc się zmagać z niezdarnymi paluchami mojego robota ani tajemniczą manetką. Problem został całkowicie

przerzucony na robota, a właściwie jego twórców: ruchy człowieka trzeba bowiem „zmapować” na ruchy robota. W idealnym przypadku musiałyby więc mieć on dokładnie taką samą anatomię, co człowiek. W przeciwnym razie doszłoby do niezręcznej sytuacji, w której biedny robot po prostu nie potrafiłby zgiąć nadgarstka tak jak ja i patrzyłby na mnie bezradnie tymi swoimi psimi oczami. Co więcej, pojawiają się tu wszystkie te paskudne „przyziemne” problemy, o których nie lubimy słuchać: musimy zadbać o dobrą widoczność i warunki oświetleniowe, nie zasłaniać żadnej części ciała i tak dalej. Zauważmy, że w dwóch pierwszych przypadkach robot otrzymuje „wprost” informację o tym, jakie ruchy są od niego wymagane. Przy uczeniu poprzez imitację musi dopiero te ruchy odczytać z kamery – bywa więc, że w czasie „sesji pokazowej” stosuje się dodatkowe czujniki położenia, montowane na przykład na dłoniach. Przy czynnościach manualnych idealnym wyjściem byłyby rękawiczki, monitorujące położenia każdego paliczka, ale bardzo jestem ciekaw, który właściciel robota domowego miałby pod ręką komplet takich rękawiczek rok po zakupie i jak by wyglądały po godzinie demonstracji kuchennych.



Widziałem to na filmie^[32]

Autorzy tego badania – związani z Uniwersytetem Kalifornijskim w Berkeley i OpenAI – nie kryją się za bardzo ze swoimi zamiarami. Postanowili nauczyć robota następujących czterech czynności: sięgania po przedmiot, popychania go w zadane miejsce, zamiatania oraz nabierania migdałów chochlą i przesypywania ich na patelnię – co każe przypuszczać, że nie jest im obojętna wizja robota pomocnika kuchennego.

Zacznijmy od prostej obserwacji: robot obserwujący człowieka przy pomocy swojej kamery tak naprawdę uczy się z filmu. Z punktu widzenia architektury algorytmu uczącego się nie ma żadnego znaczenia, czy podawany jest mu film generowany na żywo, czy też zapisany dawno temu na dysku. Zasadniczym problemem do rozwiązania jest więc „przetłumaczenie” pomiędzy filmem kręconym z perspektywy „trzeciosobowej” a ruchami robota widzianymi z jego własnego punktu widzenia.

Można to rozbić na dwa osobne zadania. Pierwszym jest zrekonstruowanie na podstawie filmu tego, co dzieje się w przestrzeni trójwymiarowej – tak, aby powstała ogólna reprezentacja „zamiatania” albo „przesuwania obiektu w miejsce X”, uniezależniona od nieistotnych detali, takich jak kolor stołu, nawet konkretna sekwencja ruchów ręką. Drugim jest późniejsze sterowanie ruchami robota tak, aby

zrealizował to samo zadanie – czyli osiągnął ten sam cel: ideą zmiataania jest umieszczenie śmieci na zmiotce, a nie zginanie kończyn w określony sposób (polecam wrócić pamięcią do architektury DIARC, gdzie pojawiła się już różnica między celem i zadaniem).

By zrealizować pierwsze zadanie, postawiono na tzw. **głębokie uczenie ze wzmacnianiem** (*deep reinforcement learning*). W ramce na stronie 24 wprowadziłem już pojęcie uczenia ze wzmacnianiem i uczenia maszynowego. Ich odmiany „głębokie” to te, w których architekturze występuje wiele warstw ukrytych; współcześnie są to również warstwy różnego typu, jak choćby te odpowiadające za „pamięć”, o czym wspomniałem w przypisach do ramki na stronie 28.

W tym przypadku były to między innymi tak zwane **konwolucyjne sieci neuronowe**, których szczególną siłą jest „uogólnianie” obrazu poprzez identyfikowanie w nim rozmaitych powtarzających wzorów – będzie o nich jeszcze mowa w ramce na stronie 117. Można by powiedzieć, że sieci takie pomijają detal, identyfikując porządek wyższego rzędu. A to właśnie o tego typu porządek chodzi, gdy próbuje się znaleźć wspólny mianownik w kilkudziesięciu filmach przedstawiających zmiataanie. Nie są istotne, przykładowo, włosy na nadgarstku osoby zmiatającej. Miłośników technologii może to zaskoczyć, ale najlepsze rezultaty osiąga się, gdy na pierwszym etapie filmom sztucznie *obniży się rozdzielczość*, tak, że wyglądają jak filmowane nawet nie starym telefonem komórkowym, ale wręcz przysłowiowym kartoflem. Dla większości zadań materiałem treningowym były filmy w rozdzielczości 48 na 48 pikseli, na których tylko z grubsza dało się dostrzec, co się dzieje. Ale o to właśnie chodziło!

Drugą sztuczką było – podobnie jak choćby w przypadku robotycznego czworonoga, który sam nauczył się chodzić (zob. ramka na s. 28) – rozpoczęcie od dużej bazy przypadków generowanych komputerowo. Dla zmiataania śmieci, przykładowo, wygenerowano 894 filmy z symulacjami komputerowymi, a liczbę rzeczywistych filmów przedstawiających zmiataających ludzi ograniczono do 85. Znowu startujemy więc w przestrzeni wirtualnej.

Algorytm generujący odpowiedni ruch robota „karmiony” był ogólną reprezentacją danej czynności, a jego zadaniem było wysyłanie do sterowników takiego sygnału, aby ostatecznie osiągnięty został zadany cel: śmieci na zmiotce, migdały na patelni.

Wyniki były niezłe, zwłaszcza w porównaniu z innymi algorytmami, na przykład takimi, które na podstawie pokazanych im materiałów szkoleniowych starały się „na sztywno” nauczyć się pewnej sekwencji ruchów, czasem wręcz sprowadzonej do osiągnięcia zadanego kąta w określonym stawie robotycznej „ręki”. Wartości procentowe skuteczności były wysokie, ale nie ma co się do nich zbyt przywiązywać – bądź co bądź, mówimy wciąż o laboratoryjnych warunkach i starannie dobranych przykładach. Liczy się sama idea.

Autorzy przytoczonego wyżej artykułu wspominają przelotnie w ostatnim akapicie o wspaniałym sposobie na wykorzystanie ich platformy: gdyby okazało się, że robot potrafi nauczyć się wykonywania pewnej czynności po prostu na podstawie odpowiednio dużego zbioru dobrze opisanych filmów, otworzyłyby to drzwi do uczenia się praktycznie wszystkiego, co można zobaczyć w internecie.

Pierwszą rzeczą, która przyszła mi do głowy, jest gotowanie – pewnie z powodu tych prażonych migdałów. W internecie można bez trudu znaleźć olbrzymią liczbę filmów prezentujących elementarne techniki (jak profesjonalne krojenie cebuli albo robienie ciasta na pizzę) i konkretne przepisy. Lista umiejętności, które można by zdobyć tą drogą, jest nieograniczona: od szydełkowania, przez malowanie figurek bitewnych i wymianę oleju silnikowego, aż po korzystanie z tokarki i odlewanie aluminium.

Mogłoby się wydawać, że może to pozwolić na stworzenie uniwersalnego superroboty. Czyż nie byłoby świetnie, gdyby nasz domowy robot pomocnik został wstępnie nakarmiony milionem filmów z internetu, nabywając dzięki temu tysiące praktycznych umiejętności? Pofantazjujmy może chwilę – zasłużyliśmy sobie po trudach przedzierania się przez opisane w tym rozdziale technologie.

Pomijając już nawet względy praktyczne (nie jest oczywiste, na ile *tak naprawdę* można się nauczyć naprawy samochodu poprzez oglądanie filmów), myślę, że tak się ostatecznie nie stanie.

Po pierwsze, producent nie ma żadnego dobrego powodu, by aż tak silnie inwestować w swoje roboty. Co więcej, jeżeli miałyby być one reklamowane jako zdolne do naprawiania aut, szydełkowania i robienia masażu, to rzeczywiście musiałyby wszystko to robić względnie przyzwoicie, bo w przeciwnym razie infolinia przegrzewałaby się od głosów oburzonych klientów, których robot przypala omlety i ma spore problemy z obsługą szlifierki kątowej. Dużo bezpieczniej byłoby sprzedawać robota z kilkoma elementarnymi umiejętnościami – nawet i nabytymi metodą analizy filmów – który byłby jednak reklamowany jako zdolny do późniejszego *nauczenia się* wszystkiego (albo, jeszcze lepiej, *uczenia się*).

Po drugie, zasoby pamięciowe i obliczeniowe nie są nieskończone. Choć popularny jest pogląd, że w przyszłości kwestie transferu danych i mocy procesorów przestaną mieć znaczenie – wizja jest taka, że dane będą przesyłane i przetwarzane w mgnieniu oka – rzeczywisty rozwój technologii nie zachęca do takiego optymizmu. Będzie o tym jeszcze mowa w rozdziale III. Nowe dane tworzone są czasem wręcz szybciej niż rośnie zdolność do ich magazynowania i przetwarzania. Na samą tylko platformę YouTube codziennie dodawane są filmy trwające łącznie 82 lata. Materiały te są ponadto z każdym rokiem coraz wyższej jakości. To dwa konkurujące ze sobą trendy: z jednej strony coraz szybszy przesył danych, coraz pojemniejsze dyski i coraz lepsze

procesory, a z drugiej strony coraz więcej użytkowników smartfonów, wykonujących coraz więcej zdjęć i nagrywających coraz więcej filmów, które są coraz częściej upubliczniane.

W miarę realistycznie można sobie natomiast wyobrazić możliwość „zlecenia” robotowi nauczenia się czegoś nowego w wolnej chwili. Robot taki „zassałby” sobie przez noc kilka tysięcy filmów z internetu – być może oferowane byłyby już gotowe „pakiety” zawierające wyłącznie użyteczne, dobrze opisane sceny? – a następnie poświęcił parę godzin na odpowiednie „podkręcenie” swoich algorytmów i próby.

Szczerze mówiąc, myślę, że to możliwość nie tylko bardziej realistyczna, ale bardziej interesująca niż robot „fabrycznie umiejący wszystko”. Po pierwsze, gdyby roboty domowe uczyły się w odpowiedzi na zapotrzebowanie użytkownika, stawałyby się z czasem coraz silniej spersonalizowane. Zauważmy, że mówimy tu nie o „wgraniu gotowego programu”, który wpada do „worka z programami”, a raczej „podtrenowaniu” złożonej sieci neuronowej przy pomocy kolejnej porcji materiału treningowego. Dwa roboty – jeden, który przez kilka tygodni byłby uczony gry na instrumentach muzycznych, oraz drugi, który w tym samym czasie uczyłby się naprawy kolejnych podzespołów samochodu – zareagowałyby później *inaczej* na tę samą lekcję gotowania.

Po drugie, wróćmy myślą do „roboty ciekawskiego” (ramka na s. 62). Wygląda na to, że lubimy roboty przejmujące inicjatywę. Gdyby więc pozostawić taką furtkę otwartą, nasz robot mógłby w wolnych chwilach spontanicznie decydować się na to – być może zainspirowany naszym „osobistym profilem elektronicznym” (zob. rozdział III.3) – by nauczyć się czegoś, co sam wybrał z katalogu tysięcy możliwości?

Na koniec krótki kontakt z rzeczywistością. Wszystko pięknie, ale na czym stoimy?

Na Międzynarodowej Konferencji Robotyki i Automatyki ICRA 2021 w chińskim mieście Xi’an, która odbyła się na przełomie maja i czerwca 2021 roku, przedstawiono wyniki Robot Household Marathon Experiment^[33]. Była to inicjatywa poddania robotów domowych „próbie ognia”: zadanie polegało na całkowicie samodzielnym przygotowaniu śniadania i posprzątaniu po nim. Przygotowany przez grupę inżynierów z niemieckiego Uniwersytetu

w Bremie robot PR2 odważnie zmierzył się z tym zadaniem i... osiągnął *prawie sukces*.

Zadanie polegało na przyniesieniu na stół miski, łyżki i kubka, a także kartonu mleka i opakowania płatków śniadaniowych, wyłącznie na podstawie poleceń głosowych typu „Mleko jest w lodówce”. Następnie robot miał nalać mleko i wysypać do niego płatki, a po skończonym śniadaniu odłożyć brudne naczynia do zmywarki, wyrzucić pusty karton po mleku do śmieci i schować płatki z powrotem do szafki. Robot miał być całkowicie autonomiczny przez cały czas trwania eksperymentu.

Obserwowanie go przy pracy^[34] przypomina nieco śledzenie trzylatka skupionego przy podobnie „poważnym” zadaniu.

Na pięć prób robotowi za każdym razem udało się „nakryć do stołu”. Miewał przy tym problemy, zwłaszcza przy wyciągnięciu łyżki z szuflady, jednak samodzielnie orientował się, że próba zakończyła się niepowodzeniem i podejmował ponownie starania, podchodząc do krnąbrnej łyżki za każdym razem nieco inaczej, aż w końcu wylądowała bezpiecznie w jego „dłoni”. Sprzątanie powiodło się tylko w trzech próbach. W trakcie jednej z nich karton mleka złośliwie wysmyknął się podczas odnoszenia go do kosza i robot nie był w stanie już go podnieść, a przy innej „palec” robota utknął pomiędzy prętami w koszu zmywarki. Zarejestrowano też przypadek, gdy niestaranie odłożone opakowanie płatków śniadaniowych zostało zgniecione podczas zamykania szuflady.

Średni czas trwania próby wynosił 90 minut.

Może mam niewygórowane oczekiwania, ale jestem naprawdę pod wrażeniem. Przecież nawet trzylatek, przy odrobinie naszej cierpliwości, może *troszkę* pomóc w domu!

Twój świat

Kolejny rozdział, kolejne marzenie.

W poprzednim fantazjowaliśmy, że w przyszłości na każde moje skiniecie będzie uniwersalny, mądry, zręczny robot. W tym pofantazjujemy, że wszystko, co mnie będzie otaczać, będzie dokładnie takie, jak sobie wymyślę, stworzone tylko dla mnie. Mój komputer i biurko, na którym go postawię, moje talerze i posiłki, które będę na nich kładł, mój dom i rower, moje ubrania, i lekarstwa – to wszystko powstanie w jednym egzemplarzu, stworzonym tylko dla mnie. Co więcej, specjalnie dla mnie powstawać też będą książki, filmy, muzyka i gry komputerowe – ściśle dopasowane do mojego gustu, spełniające wszystkie moje zachcianki.

Marzenie to jest tylko punktem krańcowym tendencji, której ślady możemy dostrzec wokół siebie już dzisiaj. Zwłaszcza w Polsce – kraju, w którym jeszcze kilkadziesiąt lat temu wszyscy chodzili w takich samych butach, jeździli takimi samymi samochodami, pili te same oranżady i oglądali te same seriale. Na całym świecie występuje jednak potężna tendencja do indywidualizacji produktów i usług, odpowiadająca na nasz rosnący ku niebiosom indywidualizm, i napędzająca go.

Z jednej strony wynika to po prostu ze stale rosnącej oferty – do wyboru mamy dziś dziesiątki tysięcy odmian butów, przypraw, zespołów muzycznych, kolorów farb i roślin doniczkowych, do tego stopnia, że podjęcie nawet najprostszej decyzji często wymaga dziś fachowej wiedzy.

O „kawę z mlekiem” może dziś poprosić tylko największy prostak. Wystarczy wejść do odpowiednio zaangażowanej kawiarni, aby być zapytany o stopień wyprażenia ziaren, kraj ich pochodzenia i grubość mielenia; następnie czeka nas wybór między espresso, doppio, lungo, macchiato, romano, americano, cappuccino, breve, flat white, frappe, frappuccino i latte. Po ustaleniu, czy interesuje nas mleko zwykłe, niskotłuszczowe, bez laktozy czy też „mleko”

migdałowe, pozostanie już tylko decyzja co do rodzaju cukru i posypki.

Fani muzyki i jej twórcy prześcigają się w kategoryzowaniu gatunków w stopniu, który nie przestaje mnie zaskakiwać. Sformułowania typu „rock”, „metal” albo „ambient” mogły cokolwiek znaczyć – i łączyć ludzi – może 30 lat temu. Już dawno nauczyłem się, czym jest „psychedelic ambient” (znany też jako „psybient”), choć dopiero parę lat temu zostałem uprzejmie poprawiony, że muzyka, której słucham, podpada raczej pod „psychedelic progressive ambient” („progpsy”), cokolwiek to znaczy. Parę miesięcy temu, wędrując sobie po obrzeżach interesujących mnie regionów^[16*], odkryłem, że na pewnym forum^[35] szybszą, bardziej transową kuzynkę psybientu, znaną jako „psychedelic trance” (czyli „psytrance”), podzielono na *siedemnaście* głównych odmian (próbka nazw: trance goa, nitzhonot, darkpsy, forest, hi-tech psybient), a po długich debatach jeszcze na kolejne. Szczególnie rozbawiła mnie zażarta dyskusja na temat konieczności wydzielenia w *forest psybient*, który operuje echem, pogłosem i ogólnie dość mrocznym klimatem (ja tam lasu nie słyszę...), pododmiany *swamp*, która, jak twierdzą koneserzy, ma w sobie nuty jednak bardziej bagniste niż leśne. Co istotne, to nie jest tylko kilku pozytywnych wariatów z drobną obsesją! Fani i twórcy, zwłaszcza niektórych gatunków, naprawdę identyfikują się poprzez tak wąsko zdefiniowane etykiety.

Skutki tego procesu są czasem zabawne, a z rzadka irytujące – jak choćby wtedy, gdy okazuje się, że następnego aspektu życia nie wolno nam przeżywać „naiwnie”, ponieważ trafiliśmy na kolejnego fanatyka, który przekonuje nas gorączkowo, abyśmy dowczyli się w sprawie sposobów dziania skarpet, odmian plastiku w kablach słuchawkowych albo wariantów pieprzu, bośmy ignoranci, a tak w ogóle to coraz więcej amatorów się pcha do zabawy. W ogólności jest to jednak wspaniały proces, któremu z całego serca kibicuję. Dzięki niemu, ostatecznie, każdy z nas żyje w nieco innym świecie.

Zauważmy przy okazji, w jak wielu dziedzinach życia przejawia się nasza niemal narcystyczna potrzeba wyjątkowości. Ot, turystyka. Wszyscy turyści zawsze chwalą się tym samym: że znaleźli jedyną taką restaurację nieuczęszczaną przez turystów, gdzie stołują się wszyscy lokalsi, a obiad ugotowała im ekwadorska babcia według przepisu babci swojej babci. Nawet zaś ci, którym nie chce się włóczyć po lokalnych miasteczkach, mogą przeżyć swoją wersję „dreszczu pioniera”. Choć gigantyczne nadmorskie hotele oferujące *all inclusive* są na całym świecie z grubsza takie same – i gdyby nawet ich

zapałonego konesera usnąć i przetransportować na drugi koniec świata, mógłby mieć poważny problem z rozeznaniem się, czy to Egipt, Seszele, Malediwy czy Puerto Rico – wartością samą w sobie stało się jednak wyprawianie do kraju „innego niż wszyscy”, oczywiście „zanim to było modne”.

Potrzebę dotarcia w coraz to nowe miejsca ilustrują nawet statystyki. W sezonie 2007/2008 największy na świecie przyrost dochodów z turystyki odnotowały: Izrael, Egipt, Turcja i Indonezja. W 2013 roku rekordowy wzrost odnotowały m.in. Kirgistan, Islandia, Bośnia, Gruzja i państwa wyspiarskie Pacyfiku, jak Tuvalu i Niue. W 2018 roku na liście „tygrysów turystyki” wiodące miejsca zajmują już Iran, Nepal, Belize, Ekwador i Uganda, dynamicznie pną się też Białoruś, a specyficzną gałęzią turystyki o rosnącej popularności są również niezbyt etyczne wycieczki do Korei Północnej. Czysto teoretycznie, w pędzie tym może oczywiście chodzić o poznanie nowych miejsc i kultur. W praktyce, milionowy turysta trafia ostatecznie do przestrzeni „globalnie uśrednionej”.

Rozwój oferty i świadomości konsumentów to jednak tylko pierwszy krok. Drugim jest coraz precyzyjniejsza kontrola nad tym, co jest konsumowane. Pomyślmy o fundamentalnej różnicy pomiędzy radiem a Spotify; pomiędzy telewizją a Netflixem. Użytkownik radia i telewizji może oczywiście wybrać sobie stację, jednak na tym kończy się jego władza decyzyjna. Użytkownik Spotify albo Netflixu może sobie samodzielnie zaprojektować kilka godzin rozrywki lub – co dzieje się coraz częściej – pozwolić to zrobić algorytmom, opierającym się na jego dotychczasowych wyborach i rozmaitych innych danych, zgromadzonych mniej lub bardziej jawnie (zob. rozdział III).

Kryje się za tym również ogólniejsza przemiana ekonomiczna: przejście od kultury posiadania do kultury dostępu. Po co mi płyta jakiegoś zespołu na półce, jeśli mogę sobie wykupić dostęp do ich muzyki? Ba, być może podoba mi się tak naprawdę na tej płycie tylko jeden utwór?

Model ten nie wszędzie się jeszcze przyjął. Od lat czekamy, przykładowo, na udaną realizację „Spotify dla książek” – usługi, która zapewniałaby równie swobodny dostęp do literatury, proponując mi na wieczór, przykładowo, jedno opowiadanie Tokarczuk z *Opowiadań bizarnych*, jedno dociekanie z *Dalszych dociekań* Borgesa i ostatni rozdział niedoczytanego poprzednio kryminału. Istnieją natomiast nieśmiałe próby przeszczerpienia tego modelu na rynek diet

pudełkowych – tak, aby każda osoba dostawała codziennie pod drzwi stworzoną tylko dla siebie kompozycję posiłków. Do rozwiązania jest jednak masa problemów prawnych, finansowych i logistycznych.

Ostatnim krokiem jest całkowite oderwanie się od modelu „wybierania z menu”, jakkolwiek rozległe by ono nie było i jakkolwiek wielka by była nasza swoboda korzystania z niego. Alternatywą jest wzięcie udziału w procesie twórczym tak, aby powstało coś tylko dla mnie. W tym rozdziale przyjrzymy się możliwościom realizacji tej wizji.

Rozdzielimy to na dwa problemy, które wymagają odmiennego podejścia: spersonalizowane rzeczy i spersonalizowane treści. Mówiąc zaś ogólniej i nieco bardziej dramatycznie, w pierwszym przypadku tworzona ma być *materia*, w drugim *informacja*.

II.1. Świat przedmiotów tylko dla ciebie

Wygląda na to, że historia zatacza ciekawe koło.

Wyobraźmy sobie prostą, półosiadłą społeczność sprzed, powiedzmy, 15 tysięcy lat. Przedmioty dostępne żyjącym w niej ludziom wytwarzane były w większości albo samodzielnie, albo przez bliskich lub dalszych znajomych. W każdej społeczności musiał być ktoś, kto potrafił produkować i naprawiać naczynia, narzędzia czy ubrania. Przedmioty te nie były więc produkowane masowo, a następnie wystawiane w sklepie, w którym kupowały je anonimowe osoby – konkretna osoba zgłaszała się z konkretnym zapotrzebowaniem: potrzebny jest kaftan dla Heńka i nóż dla Marysi. To dawało spore pole do indywidualizacji.

Wraz z rozwojem osiadłego trybu życia oraz pojawianiem się nowych materiałów i technologii szybko okazało się jednak, że model ten ma poważne ograniczenia. Czy w każdej społeczności rzeczywiście musi znaleźć się ktoś umiejący obrabiać obsydian, wiercić dziurki w kamykach, wytapiać miedź? Co więcej, osoby wydobywające owe cenne minerały, kamienie i metale miały żywy interes w tym, żeby eksportować nie sam surowiec, co raczej wytworzone z niego produkty. Na długo przed pierwszymi zaawansowanymi cywilizacjami Sumeru, Egiptu czy doliny Indusu świat opłatały już rozliczne szlaki handlowe, którymi wędrowały narzędzia, broń i ozdoby. Wraz z rozwojem cywilizacji coraz silniejsza stawała się standaryzacja produkcji. Przez setki i tysiące lat, jeszcze przed epoką produkcji maszynowej, statkami i karawanami kursowały

tysiące identycznych egzemplarzy naczyń, narzędzi, ubrań i ozdób. Rewolucja przemysłowa XVIII i XIX wieku stanowiła tylko tak naprawdę puentę tego ogólnego procesu cywilizacyjnego – ponieważ maszyna wytwarza przedmioty *naprawdę niemal identyczne* – zaś otwarcie się świata dla handlu i wpływów kulturowych sprawiło ostatecznie, że ja i mój równolatek z papuaskiego lasu tropikalnego możemy nosić dokładnie taką samą koszulkę z okładką płyty Pink Floyd. Mottem dla tej fazy rozwoju cywilizacji są słowa Henry’ego Forda, ojca chrzestnego nie tylko samochodu osobowego, ale i produkcji masowej jako takiej: „Twoje auto może mieć dowolny kolor, byleby to był czarny”.

Filmy i seriale science-fiction bardzo lubią przedstawiać mieszkańców przyszłości – i członków pozaziemskich ras inteligentnych – w jednakowych monochromatycznych skafandrach. Wygląda jednak na to, że ów proces unifikacji zaczyna się powolutku odwracać. Stajemy się coraz bardziej wybrednymi konsumentami i chcemy dostać dokładnie to, co sobie wymyśliliśmy, skrojone ściśle dla naszych potrzeb.

Niektóre tego skutki opisałem już wyżej. Teraz zastanówmy się, jak można by wykonać ten ostatni krok, dzięki któremu – znów, jak w wyidealizowanych społecznościach tradycyjnych – przedmioty, z których korzystam, tworzone by były tylko dla mnie.

Jednym jest faktyczny **powrót do rzemiosła tradycyjnego** – realny i interesujący trend, który może nabrać rozpędu, zwłaszcza kiedy sprzymierzy się go z ideami takimi jak ucieczka do natury, *slow food* i żywność ekologiczna oraz powrót do małych wspólnot opartych na wymianie dóbr i usług, z lekką nutą anarchokomunizmu w tle. Niewykluczone, że w XXI wieku więcej osób będzie korzystało z usług kowali, tkaczy i kapeluszników niż w wieku XX. Ponieważ jednak trend ten nie wiąże się z jakimś istotnym postępowaniem technologicznym (a jest wręcz, w najlepszym sensie, technologicznym „krokiem wstecz”), nie będę przyglądał mu się bliżej.

Drugą ciekawą furtką otwierają rosnące możliwości produkowania coraz bardziej zaawansowanych technologicznie przedmiotów w warunkach domowych. Rośnie zainteresowanie drukiem 3D, domową syntezą chemiczną, produkcją własnej elektroniki, spawaniem, odlewaniem i toczeniem, a urządzenia dające taką możliwość stają się coraz tańsze i coraz prostsze w obsłudze.

Przydomowa **fabryczka w garażu** to bardzo ciekawa koncepcja i to jej poświęcona będzie pozostała część tego rozdziału.

Warto jednak wspomnieć pokrótce o jeszcze innej możliwości. Nie zapowiada się na to, że wielkie siły rynkowe zrezygnują z walki i oddadzą segment rynku produktów „hiperindywidualizowanych” podmiejskim komunom hipsterów i garażowym drukarzom ekspresów do kawy. Dziś jest to już jeden z trendów marketingowych: **produkcja spersonalizowana** (*personalized manufacturing*), której towarzyszy spersonalizowany marketing. Nike oferuje swoim klientom Nike By You: narzędzie do samodzielnego projektowania butów, które mniej więcej miesiąc później trafią pod nasze drzwi. Możliwości indywidualizacji są oczywiście ograniczone. Po wybraniu jednego z dostępnych modeli buta (w momencie pisania tego tekstu było ich 24) otrzymujemy możliwość ustalenia barwy dla wszystkich pól kolorystycznych, a następnie wpisania 7-literowego tekstu, który zostanie naszyty na język. Dla wybranego przeze mnie na próbę modelu oznaczało to podjęcie 11 decyzji kolorystycznych, co łącznie daje, plus minus, 100 miliardów możliwości. Oczywiście, za każdym razem będzie to wciąż ten sam but, tylko inaczej pokolorowany: pamiętam dość podobną procedurę z dzieciństwa, kiedy to kupione na targu trampki barwiło się pisakiem; niektórzy odważni domalowywali też paski Adidasa albo „znaczek najki”.

Podobny poziom indywidualizacji oferują dziś również choćby producenci samochodów: Nissan chwali się, że klient może samodzielnie wybrać sobie między innymi typ silnika i wygląd wnętrza, aby powstało „auto jedyne w swoim rodzaju”. W praktyce procedura jest jednak równie mało ekscytująca, co kolorowanie butów, a Nissan tym się najwyżej różni od swoich konkurentów, że silniej od nich podpiął się pod trend marketingowy i reklamuje się hasłem „Build Your Nissan”.

Wygląda na to, że reklamy są póki co znacznie bardziej kolorowe od rzeczywistości. Nie bagatelizujmy jednak samego trendu. Rozważa się czasem przekonstruowanie fabryk tak, aby z linii produkcyjnej schodziły tysiące różniących się od siebie produktów. Idea ta ma też swoją nazwę: **indywidualizowana produkcja masowa** (*individualized mass production*). Od strony „sprzętowej” sąsiaduje ona z koncepcjami takimi jak „sprytnie fabryki” (*smart factories*) i Przemysł 4.0. Czysto teoretycznie, wymagania takie mogłyby spełnić coraz bardziej elastyczne i „inteligentne” roboty przemysłowe, bardzo dobrze ze sobą skomunikowane, pracujące zgodnie z procesami produkcyjnymi generowanymi na bieżąco przez algorytmy. Docelowo, odpowiednio

elastyczna, modułarna linia produkcyjna miałyby tworzyć od zera tysiące telefonów komórkowych, laptopów albo samochodów, każdy z myślą o konkretnym kliencie. Jest to jednak kierunek ściśle eksperymentalny i nie ma jeszcze na świecie żadnej naprawdę zarabiającej na siebie fabryki tego typu. Co więcej, istnieją poważne wątpliwości, na ile konsument jest tak naprawdę w stanie cokolwiek zaprojektować: o ile nieprzemysłany wybór tapicerki sprawi w najgorszym razie, że auto będzie paskudne, to przecież parametry projektowe smartfona czy samochodu wyłaniają się na drodze wieloletnich procesów optymalizacyjnych. Krótko mówiąc, może się okazać, że każdy proces produkcyjny wystarczająco złożony, żeby domagał się linii produkcyjnej, może być też zbyt złożony, żeby umożliwił indywidualizację. To trudny temat.

My tymczasem przyjrzymy się – mówiąc starym, dobrym językiem marksistowskiej socjologii – jak mogłoby wyglądać przekazanie środków produkcji masom. To coś, co nie tylko ma spory potencjał, ale w dodatku już się dzieje. Wiatr rewolucji!

Zacniemy od wizji szczególnie kuszącej: czyż nie byłoby pięknie, aby w moim własnym garażu znajdował się sprzęt zdolny do „wyplucia” wszystkiego, co sobie tylko wymarzę – z taką samą łatwością, z jaką drukarka zapełnia kartkę czymkolwiek, co tylko narysuję na ekranie komputera? Ma to w sobie powab, którego nie uświadczymy, zamawiając coś wyjątkowego z chińskiej fabryki, nawet reklamowanej słowem „smart”. Zresztą, przecież sprzęt taki mógłby również służyć do uwalniania kreatywności i zabawy, a nie tylko do generowania przydatnych produktów.

Fantazja unosi nas szybko. Przecież wystarczyłoby stworzyć coś, co porządkuje atomy w zadany kształt, prawda? Ech, niby tak. Tego typu technologii nie ma jednak jeszcze nie tylko w załączku, ale nawet na poziomie „urządzeń z grubsza wyobraźalnych”. Musimy myśleć ostrożniej, a nasze marzenie realizować po trochu.





Druk 3D

Drukowanie 3D jest już jedną z tych technologii, które weszły pod strzechy. Można to poznać choćby po tym, że doczekały się już swojej kategorii na Allegro, a najprostsze drukarki dostępne są za przysłowiowe „parę stów”. Mało tego, technologia ta jest już podstawą dla innych technologii – wiele pionierskich badań i eksperymentów omawianych w tej książce opiera się na prototypach drukowanych właśnie tą metodą. Jeżeli więc chcemy bawić się w łączenie kropek, to z węzła „drukarki 3D” prowadzą kreski do niemal wszystkich innych kropek na naszej mapie przyszłości technologii.

Tym bardziej warto wrócić do podstaw. Chociaż najbardziej znaną formą drukarek 3D są chyba te, w których z ruchomej głowicy wyciskany jest warstwami stopiony plastik, istnieją dziesiątki metod, za pomocą których można zrealizować ten zasadniczy cel, jakim jest wytworzenie trójwymiarowego obiektu o zadanym kształcie poprzez automatyczne budowanie go „po kawałeczku”. W pierwszej chwili można się zagubić w mnogości tych technologii, zwłaszcza że nazwy ich są wyjątkowo nieprzyjazne dla ucha i zwykle ukrywa się je za tajemniczymi trzyliterowymi skrótami, jak FFF, FDM, SLA, SLM, SLS, EBM czy CAL.

Nie bójmy się jednak tych technologii – których znaczenie będzie w kolejnych latach tylko rosło, a których zasada działania jest zwykle względnie prosta. Mówiąc w przenośni i dosłownie: to nie fizyka kwantowa. Skupmy się może na kilku głównych klasach.

Czysto historycznie, pierwsza faktycznie zrealizowana technologia opierała się na **fotopolimeryzacji**: utwardzaniu światłem płynnego polimeru. W 1981 roku japoński inżynier Hideo Kodama opisał w jednym artykule^[36] aż trzy urządzenia, przy pomocy których już rok wcześniej wytworzył pierwsze trójwymiarowe modele. Ogólna zasada ich działania jest następująca: istnieją płynne polimery twardniejące pod wpływem światła – w kontekście druku 3D mówi się na nie dziś „żywice”. Kodama zauważył, że płyn taki po naświetleniu światłem ultrafioletowym tworzy na powierzchni twardą warstewkę, trochę jak zastygła „kra” wosku. Wyobraźmy sobie następnie, że wyłączamy wówczas lampę UV, powstałą krę zanurzamy nieznacznie pod powierzchnię płynu – na głębokość odpowiadającą grubości kry – i ponownie włączamy lampę. Po odpowiednim dobraniu grubości i czasu naświetlania nowo twardniejąca kra scali się ze starą, tworząc solidny blok polimerowy. Gdyby zaś na lampę ultrafioletową nałożyć maskę, blokującą promieniowanie w ściśle określonych miejscach, to „kra” mogłaby mieć dowolny kształt. Kodama stworzył serię takich masek, odpowiadającą kolejnym poziomym przekrojom przez wymyślony obiekt 3D, a następnie wykonał w ten sposób pierwszy przedmiot wydrukowany na „drukarcie 3D” – był to model domu o wymiarach 7×5×5 cm zawierający 27 warstw. W środku

znajdowały się ściany, meble i schody, a jego „wydrukowanie” zajęło 4,5 godziny. W tym samym artykule opisane jest też urządzenie, w którym nie ma masek, zaś płyn naświetlany jest lokalnie przez wędrującą nad nim głowicę z laserem UV.

Trzy lata później Amerykanin Chuck Hull ukuł dla tej metody nazwę **stereolitografia (SLA)** i złożył patent na oparty na niej proces produkcyjny (Kodama i konkurencyjna grupa francuska nie zdążyli ze swoimi patentami). W 1986 roku Hull założył do dziś istniejącą firmę 3D Systems, często nazywaną pierwszą na świecie firmą specjalizującą się w druku 3D. Istnieją dwa podstawowe sposoby na to, jak można wykorzystać fotopolimeryzację do tworzenia obiektów trójwymiarowych: opisana wyżej metoda „plasterkowa” albo „cała naraz”: **vat photopolymerization**, czyli „fotopolimeryzacja w zbiorniku” (wszyscy są tak przyzwyczajeni do trzyliterowych skrótowców, że metodę tę określa się czasem jako VAT). Przykładową realizacją VAT jest xolografia – poświęcam jej osobną ramkę niżej. W 2015 roku opisano też^[37] ciekawą pośrednią technikę o nazwie CLIP, w której obiekt jest w sposób ciągły „wyciągany” z kąpieli polimerowej: dzięki temu nie składa się ostatecznie z osobnych plasterków, a jego produkcja jest znacznie szybsza. Filmy przedstawiające wytwarzanie figurek metodą CLIP robią piorunujące wrażenie: w kilka minut z wanienki z płynem wyciągnięta zostaje piękna gładka figurka, z detalami o rozmiarach liczonych już nie w milimetrach, a mikrometrach.

Niedługo po opisaniu stereolitografii pojawił się jej „proszkowy” odpowiednik. Wyobraźmy sobie, że zamiast płynu, który miejscowo utwardzamy, posługujemy się proszkiem, który miejscowo spiekamy. Podstawowa metoda jest dość prosta: płytki basenik pokrywamy cienką warstwą proszku – mogą to być kuleczki metalowe, ceramiczne, plastikowe, a nawet szklane^[38] – a następnie w niektórych miejscach silnie go ogrzewamy, na przykład laserem. Granulki stapiają się ze sobą, a my наносimy kolejną warstewkę proszku – i tak dalej. Ogólna technologia nosi nazwę Powder Bed Fusion (czyli dosłownie „łączenie w złożu proszkowym”), a jego najbardziej znaną realizacją jest SLS – **Selective Laser Sintering** (czyli dosłownie „wybiórcze spiekanie laserem”). Jej „podkrecona” wersja, w której granulki nie są tylko „nadtapiane” i spiekane brzegami, ale ulegają całkowitemu stopieniu, to Selective Laser Melting (SLM) – ach, kolejny skrót.

Ciekawostka dla fanów „łączenia kropek”: technologia SLS spotkała się z dużym zainteresowaniem przemysłu farmaceutycznego (czy to dlatego, że farmaceuci naturalnie „lubią się” z proszkami?). Jednym z kierunków badań są zindywidualizowane kapsułki z lekami, projektowane tak, aby uwalnianie substancji czynnej dokonywało się w ściśle zaprogramowanym tempie, na przykład z myślą o konkretnym pacjencie i jego metabolizmie. W artykule przeglądowym z 2020 roku^[39] opisano kilkanaście eksperymentów tego typu, w których z odpowiedniego materiału, ulegającego rozpuszczeniu w organizmie, drukuje się metodą SLS kapsułkę uwalniającą lek w praktycznie dowolnym tempie, na przykład kolejnymi dawkami co 6 godzin, w całości po 8 godzinach albo z powolnym uwalnianiem rozciągniętym na 12 godzin.

Najbardziej dziś rozpowszechniona technika druku 3D nosi nazwę **Fused Filament Fabrication (FFF)** – „produkowanie poprzez stapianie filamentu”^[17*]. Jest spora szansa, że widzieliście już taką drukarkę: nad stołem roboczym przesuwają się głowica, nagrzewająca podawany do niej długi plastikowy drucik („filament”), który stapia się i przykleja do podłoża. Ideę zrozumie szybko każdy, kto bawił się kiedyś pistoletem z pałeczkami do klejania na gorąco albo oglądał w dzieciństwie bajkę o zaczarowanym ołówku^[18*].

To popularna, tania technologia, nic więc dziwnego, że jest rozwijana w rozmaitych kierunkach. Jedni idą w rozdzielczość. W 2017 roku opisano^[40] proces „drukowania” miedzią, przypominający co do zasady FDM, w którym głowica pozostawiała za sobą ślad o grubości około 1 mikrometra (μm), czyli jednej tysięcznej milimetra. Autorzy bawili się między innymi w pisanie literkami o wysokości około 20 μm . Inni eksperymentują z drukarkami zdolnymi do jednoczesnego deponowania wielu materiałów: do głowicy podawanych jest kilka różnych filamentów. Przykładowo, w 2019 roku opisano^[41] drukarkę tworzącą przedmioty plastikowo-metalowe.

Gdyby zrobić „krok wstecz”, technologię FFF można byłoby opisać nieco ogólniej jako „dostarczanie jakiegoś materiału z podajnika w ściśle określone miejsce, a następnie unieruchomienie go tam”. Jeżeli zdefiniujemy problem tak ogólnie, to utworzą nam się inne „klapki”.

Ot, technologia, która jest już tak bardzo podobna do zwykłej biurowej drukarki atramentowej, że bardziej nie można: **3D Inkjet Printing**, czyli... „drukowanie atramentowe 3D”. Pionierzy tej metody stosowali zwykłe drukarki atramentowe, w których zamiast tuszu używano jakiegoś innego płynu: na tyle rzadkiego, aby dało się go przecisnąć przez dysze, ale na tyle lepkiego, aby dało się go natryskiwać warstwami. Ostatecznie okazało się, że funkcjonalne dysze trzeba będzie jednak wyprodukować osobno – dzięki temu technologia ta ostatecznie się usamodzielniała. Szczególnie dobrze sprawdza się wtedy, kiedy pracujemy z materiałami wilgotnymi albo lubiącymi wodę, jak... żywe komórki.

Jedną z pierwszych propozycji zastosowania tej metody było drukowanie organów – cóż bowiem „prostszego” niż wyposażyć drukarkę 3D w kilkadziesiąt zbiorniczków zawierających zawiesiny z różnymi typami komórek, a następnie odtworzyć pracowicie, warstwa po warstwie, strukturę nerki albo wątroby. Praktyka okazała się nieprosta, jednak biotechnolodzy znaleźli zastosowania tej technologii: dzisiaj^[42] drukuje się w ten sposób na przykład rusztowania do hodowli komórkowych (zob. rozdział IV.1.1) albo miniorgany, na których bada się rozwój chorób (zob. rozdział IV.2.1).

W zupełnie innej skali ta sama zasada ogólna doprowadziła do idei „**drukarki architektonicznej**” (tak, i na to mamy paskudny skrót: 3D Construction Printing, czyli 3DCP). A co, gdyby drukarkę dziesięciokrotnie powiększyć, a materiałem podawanym do ruchomej głowicy był szybko zastygający beton (albo jakiś jego stosowny odpowiednik)? Wizjonerzy szybko wpadli też na pomysł, aby przyszłe kolonie na Księżycu albo na Marsie budowane były na miejscu z lokalnego gruzu („regolitu”) przez „kombajny”, pobierające go i rozdrabniające, a następnie – po nasączeniu odpowiednim „klejem” natryskujące go do postaci budynków. Póki co technologie te są w fazie bardzo wczesnych prototypów – głównie dlatego, że sam „szkielet” budynku to w gruncie rzeczy i tak najprostszy element czegoś, co ma się stać funkcjonalnym domem.

Jak daleko można się posunąć z drukiem 3D? Nikt nie wie na pewno: jak widać z powyższego zestawienia, nie jest to tak naprawdę jedna konkretna technika, a raczej cała gałąź technologii produkcyjnej. Co więcej, niemal zawsze są to urządzenia niewielkie, „mieszczące się na stole”, względnie nieskomplikowane i doskonale rokujące jako wyposażenie garażu odpowiednio zmotywowanego hobbyisty.

Kluczowy jest też unoszący się nad tą technologią „opar otwartości”. Bardzo często twórcy i zwolennicy drukowania 3D to jednocześnie ludzie aktywnie działający na rzecz **otwartej nauki i technologii** – modele 3D i schematy urządzeń są powszechnie publikowane na licencjach otwartych (typu Creative Commons), co oznacza, że każdy może z nich korzystać za darmo bez żadnych ograniczeń. Ba, już w 2005 roku pojawił się pierwszy pomysł na drukarkę 3D dającą się... wydrukować na drukarce 3D. Nazwa projektu to RepRap. Choć ostatecznie nie da się wydrukować *wszystkich* elementów funkcjonalnej drukarki 3D, istnieją już modele – jak Snappy – w których drukuje się *większość* elementów. Na stronie internetowej projektu^[43] można znaleźć pełną listę niezbędnych części, modele 3D dla wszystkich elementów drukowalnych oraz przystępną instrukcję montażu, przypominającą nieco książeczki dodawane do klocków Lego.

Fraza „open source” jest niemalże mottem dla „ruchu makerów”^[19*] („wytwórców”) – idea jest taka, żeby jak najwięcej osób samodzielnie tworzyło: jak najwięcej rzeczy, jak najlepszej jakości, jak najłatwiej i jak najtaniej. Istnieją już grupy zapaleńców i organizacje udostępniające za darmo schematy konstrukcyjne, wraz z modelami dla drukarek 3D, dla systemów oczyszczania wody (AguaClara), laptopów (VIA OpenBook) i turbin wiatrowych (Zoetrope). Wciąż niezrealizowane jest marzenie stworzenia w pełni „otwartego” modelu samochodu – choć działa kilka grup żywo nad tym pracujących. Istnieje już natomiast funkcjonalna **broń open source**, stworzona z myślą o dostępnych komercyjnie drukarkach 3D, sprawiająca zresztą niemały problem prawny na całym świecie. Początkowo były to proste i mało precyzyjne „jednostrzałowce”, ale w październiku 2020 roku opisano pierwszy drukowany karabin maszynowy, Scz0rpion (choć do jego zbudowania wciąż potrzebne są niektóre części z prawdziwego czeskiego karabinu Skorpion).

Idźmy jednak dalej. Drukowanie 3D przy pomocy jakiegoś materiału to oczywiście ekscytująca wizja. Co jednak, gdybyśmy chcieli zejść nieco „głębiej” i zacząć kontrolować wedle uznania samą materię?





Automatyczna synteza chemiczna^[44]

Na listę osób, które mogą stać się w przyszłości marką samą w sobie i hasłem wywoławczym – jakim dziś jest choćby Elon Musk – ma szansę trafić Lee Cronin, brytyjski chemik, który od lat z wielkim powodzeniem zajmuje się niemal wszystkim, co ekscytujące w chemii. Pochodzenie życia, chemia supramolekularna, samoorganizacja w nanotechnologii... Jego grupa badawcza przy Uniwersytecie w Glasgow, nosząca po prostu nazwę Cronin Group, robi naprawdę pionierskie badania we wszystkich tych dziedzinach. Jednym z koników Cronina, któremu przyjrzymy się bliżej, jest automatyczna synteza chemiczna.

„Standardowa” synteza chemiczna (gdzie „synteza” to po prostu wytwarzanie jakiegoś związku chemicznego) wykonywana jest ręcznie: nawet jeżeli poszczególne kroki, jak odwirowanie związku A z mieszaniny A+B, przeprowadzane są przez automat, to całą procedurę przeprowadza człowiek: od wybrania i odmierzenia reagentów i rozpuszczalników, przez nastawienie urządzeń, aż po „głupie” noszenie probówek z miejsca na miejsce. Proces ten udało się oczywiście zautomatyzować na skalę przemysłową, jednak pojedynczy człowiek, pragnący wytworzyć jakiś związek chemiczny, musi zakasać rękawy i trochę się namachać. Idea automatycznej syntezy chemicznej jest taka, aby człowiek wykonał tylko pierwszy krok: zapewnił maszynie niezbędne związki początkowe („substraty”), a potem tylko wcisnął guzik i bachnął się w leżaku z książką.

Jedna z możliwości jest taka, że tworzymy mobilnego robota z chwytynymi kończynami i wpuszczamy go do profesjonalnego laboratorium – trochę w myśl tego, co opisywałem w rozdziale I. To jednak tak naprawdę półśrodek, w dodatku mało interesujący dla majsterkowiczów-amatorów. Prawdziwym przebojem byłoby stworzone od zera „automatyczne laboratorium”, od początku zaprojektowane w taki sposób, aby obsługiwał je komputer, a nie człowiek.

Jest to żywy przedmiot badań wielu grup z całego świata i co parę miesięcy publikowane są nowe modele „chemicznej maszynki”, jak choćby sprytny model „radialny” opisany w marcu 2020 roku przez grupę z niemieckiego Instytutu Maxa Plancka^[45], wyposażony w centralnie położony „węzeł”, do którego można podpinać dowolną liczbę modułów, a następnie korzystać z nich w dowolnej kolejności. W styczniu 2019 roku grupa Cronina opisała jednak cały pobudzający wyobraźnię „pakiet idei”, w której sama „maszyna chemiczna” o intrygującej nazwie „chemputer” to tylko wierzchołek góry lodowej.

Idea tego urządzenia – podobnie jak i samej automatycznej syntezy – jest prosta: to układ zbiorników oraz urządzeń laboratoryjnych połączonych ze sobą przewodami, wyposażony w system pomp. Związki przepompowywane są w określonej kolejności do odpowiednich urządzeń, których nastawienie również dokonuje się automatycznie. W „chemputerze” Cronina znajdują się tylko cztery urządzenia: reaktor, filtr, rozdzielacz i ewaporator (wyparka). Reaktor to zbiornik, w którym zachodzi pożądana reakcja chemiczna, co w najprostszym przypadku wymaga po prostu wpompowania do niego dwóch związków i ścisłego kontrolowania temperatury uzyskanej mieszaniny przez zadany czas. Filtr służy do oddzielania

cząstek stałych z cieczy. Rozdzielacz to urządzenie, w którym do roztworu dodaje się odpowiednio dobrany rozpuszczalnik, który po solidnym wstrząśnięciu „przechwytuje” określone składniki tego roztworu, a po uspokojeniu się zbiornika rozdziela się, albo wypływając na powierzchnię, albo tonąc. Z tak rozdzielonej mieszaniny można następnie „odciągnąć” pożądany składnik (podobną procedurę można przeprowadzić w kuchni, dodając nieco tłuszczu do sosu, a następnie zbierając go po wymieszaniu z powierzchni – w ten sposób usuniemy z sosu wszystkie składniki rozpuszczające się w tłuszczach, a więc, niestety, sporą część smaku i aromatu). Ewaporator to z kolei urządzenie przeprowadzające kontrolowane podgrzewanie roztworu i zbierające kolejno odparowujące związki (tak z kolei usuwa się z potrawy alkohol, który wrze w 78°C). Pompy, podgrzewacze, ochładzacze i sterowniki wszystkich urządzeń znajdują się pod kontrolą centralnego komputera, który realizuje zadany „program chemiczny”. I... tyle.

Już to samo w sobie pobudza wyobraźnię, ale to nie koniec. Twórcy komputerów chemicznych od lat narzekają, że szlaki syntezy („przepisy”) publikowane w fachowych czasopismach chemicznych, choć na pozór wydają się precyzyjne, w rzeczywistości takie nie są. Stosuje się określenia typu „dobrze wymieszać”, „solidnie wstrząsnąć”, „dodawać, aż mieszanina stanie się przezroczysta” albo „ochładzać do wytrącenia się żółtawych kryształków”. Doświadczony chemik rozumie, co te określenia znaczą w praktyce i na bieżąco monitoruje swoje kolby, odpowiednio reagując na zachodzące w nich przemiany – nawet jednak dobry chemik może czasem potrzebować do tego kilku prób albo doczytania literatury. Tu jednak chcemy cały ten proces zautomatyzować, co oznacza konieczność ścisłej standaryzacji: „przepisy” trzeba tak naprawdę przetłumaczyć na sekwencję konkretnych poleceń dla pomp i podgrzewaczy. Autorzy pracy opisują więc w pracy formalny język programowania procedur chemicznych, przekształcając tym samym swoje urządzenie w coś w stylu „komputera chemicznego” (stąd „chemputer”). Dokumentacja tego języka została opublikowana na licencji otwartej tak, aby każdy zainteresowany mógł z niego korzystać, a – wspominałem już, że Cronin to niepoprawny marzyciel? – szlaki syntezy nowych związków były w przyszłości już zawsze publikowane w postaci kodu komputerowego, a nie „mętnych” opisów w języku angielskim – równie przestarzałym, co łacina.

Nawet i to jednak nie wszystko. Autorzy opisują kompletną „ścieżkę przepływu informacji”, od szlaku syntezy opublikowanego w artykule naukowym – po prostu tak, jak został opisany – aż po „kod syntezy” zrozumiąły przez „chemputer”. Najtrudniejszy jest tu pierwszy krok: stworzenie listy instrukcji na podstawie treści artykułu naukowego, napisanego w języku naturalnym, z całym właściwym dla niego brakiem precyzji. Rozumienie języka naturalnego i wyciąganie konkretów z tekstu pisanego „po ludzku” to jednak coś, co komputery robią coraz lepiej. Nie powinno nas więc chyba dziwić, że w lipcu 2020 roku grupa naukowców ze szwajcarskiej gałęzi IBM Research opisała algorytm, który ma wykonywać właśnie to niełatwe zadanie – tłumaczenie szlaków syntezy z literatury naukowej na konkretne sekwencje czynności – automatycznie, dzięki mocy głębokiego uczenia maszynowego^[46]. Gdyby plan ten udało się zrealizować, można byłoby za jednym zamachem otrzymać zbiór zrozumiałych dla „chemputera” przepisów na praktycznie dowolny związek chemiczny.

Konsekwencje posiadania tego typu „fabryczki chemicznej” są potężne. Autorzy cytowanej wyżej pracy tymczasem bez żadnego

zażenowania wychodzą naprzeciw myślom chemików amatorów. Przede wszystkim celowo skonstruowali swój „chemputer” z powszechnie dostępnych części i urządzeń. Co więcej, na stronie projektu Chemify^[47] znajdują się obecnie szczegółowe instrukcje pozwalające skonstruować „chemputer”, włącznie z listą części, z dokładnością do konkretnych numerów katalogowych (przewidziano też miejsce dla listy modeli dla drukarek 3D), oraz lista gotowych „przepisów” do ściągnięcia. Całkowity koszt wyprodukowania „chemputera” szacowany jest tam na mniej niż 10 tysięcy dolarów.

Co więcej, jednym z trzech związków, na których postanowili przetestować urządzenie, był sildenafil – znany nieco lepiej pod nazwą handlową Viagra. Autorzy artykułu chwalać się, że ich „chemputer” po 102 godzinach intensywnej, całkowicie samodzielnej pracy wyprodukował porcję czystego sildenafilu, w ostatnim kroku nawet wysuszoną (poprzez przedmuchiwanie argonem w temperaturze 50°C) dla wygody użytkownika. Przepis ten każdy może pobrać ze strony projektu Chemify. Chyba tylko przyzwoitość powstrzymała ich przed pochwaleniem się, jak ich sprzęt radzi sobie z przepisem na amfetaminę, a brak pancernego szkła przed produkcją TNT.

Lekarstwa, narkotyki i materiały wybuchowe to trzy główne klasy substancji kontrolowanych – tj. takich, których nie wolno wytwarzać bez specjalnego zezwolenia. Upowszechnienie się „przydomowych fabryczek” mogłoby poważnie wpłynąć na rynek tych substancji. Nie można realnie spodziewać się skutecznej kontroli na poziomie sprzętu (którego elementy otwarcie „reklamuje się” w powyższym artykule jako *off-the-shelf*, czyli dostępne w zwykłym obiegu komercyjnym) i „przepisów” (które sprowadzają się do krótkiego pliku z kodem). Jedyne nadzieje to kontrola nad półproduktami – co dzieje się już zresztą dzisiaj.

Na pozór nieinteresujący bezbarwny oleisty płyn o nazwie fenyloaceton, który nie jest ani wybuchowy, ani psychoaktywny, należy do substancji ściśle kontrolowanych tylko dlatego, że stanowi substrat w popularnej ścieżce syntezy amfetaminy. Oczywiście, można pójść o krok wstecz i podjąć próbę wyprodukowania fenyloacetonu z czegoś legalnego, albo przynajmniej mniej kontrolowanego. Domorośli chemicy i ustawodawcy wędrują jednak po tych samych ścieżkach i substancją kontrolowaną jest też choćby kwas fenylooctowy: w temperaturze pokojowej biaława masa

o miodowym aromacie, z której nie ma zbyt wiele pożytku – no może z wyjątkiem tego, że można z niego uzyskać fenylacetone.

Zupełnie inną ścieżką jest oczywiście ekstrakcja z roślin, zawierających często w sobie koktajl najrozmaitszych fascynujących związków. Ich uzyskanie w postaci czystej ma jednak w sobie często więcej ze sztuki niż „eleganckiej” chemii laboratoryjnej. Rośliny produkują zwykle złożone mieszaniny blisko ze sobą spokrewnionych związków, których proporcje lubią być uzależnione od szczepu, czasu zbioru czy nawet pogody, a ich rozdzielenie nie jest proste.

Czysto teoretycznie, można by oczywiście wyobrazić sobie sprowadzanie szlaków syntezy do poziomu prostych cząsteczek, jeśli nie wręcz atomów. W idealnym przypadku użytkownik napełniałby więc wielki zbiornik gotowym „koktajlem” przypominającym sól fizjologiczną, zawierającym wodę z paroma kluczowymi związkami wyposażonymi w łatwe do „odcinania” główne grupy funkcyjne. To jednak z kolei prowadziłoby do komplikowania się urządzenia, które musiałoby albo zawierać pokaźną bibliotekę enzymów, albo przeprowadzać reakcje w wysokich temperaturach i ciśnieniach, reakcje fotochemiczne i tym podobne. Krótko mówiąc, garażowa szafeczka chemiczna nagle stałaby się monstrualnym kombinatem, diablo kosztownym w budowie i utrzymaniu.

Kompromisy, zawsze te kompromisy...

Zastosowania nielegalne i półlegalne rzeczywiście narzucają się jako pierwsze, ale nie ograniczamy się do nich. Jedną z możliwości, o której warto wspomnieć w duchu „łączenia kropek”, to synteza materiałów dla drukarek 3D, jak żywice. A co z pożywieniem? W literackich wizjach przyszłości bardzo lubi się pojawiać „obrzydliwa zupka odżywcza” – koktajl tłuszczów, aminokwasów, węglowodanów i witamin, zaspokajający wszystkie elementarne potrzeby ludzkiego organizmu, choćby w trakcie długich podróży kosmicznych^[20*], ale drukarka 3D wcale nie musi generować jednolitej papki.

Co więcej, ostatecznie może się jednak okazać, że nawet w drodze na Marsa nie musimy być skazani ani na pożywne zupki, ani na kosztowne transporty z powierzchni Ziemi. Wytworzenie świeżego mięsa wcale nie jest takie trudne – i nie wymaga udziału zwierząt.





Rolnictwo komórkowe

Termin rolnictwo komórkowe (*cellular agriculture*) jeszcze się nie upowszechnił, ale prawdopodobnie słyszeliście już o jego najsłynniejszym przykładzie, jakim jest **mięso hodowane *in vitro***. W rzeczywistości jest to jednak gałąź technologii, której ideę można by zdefiniować dużo ogólniej – jako wytwarzanie produktów roślinnych i zwierzęcych, ale bez powoływania do życia tych organizmów, wyłącznie w oparciu o ich komórki i tkanki hodowane w warunkach laboratoryjnych. W praktyce chodzi zaś niemal wyłącznie o produkty odzwierzęce, takie jak jaja, mleko, sery i mięso oraz skórę.

Podstawą dla rolnictwa komórkowego są zdobycze biotechnologii w zakresie dotyczącym komórek macierzystych i inżynierii tkankowej. Typowy plan hodowli mięsa laboratoryjnego rozpoczyna się od pobrania od danego zwierzęcia – kury, krowy, świni – komórek macierzystych (zob. ramkę na s. 173). Komórki te są następnie hodowane w takich warunkach, aby ich kolonia jak najbardziej upodobniła się do mięśnia, zgodnie z podstawowymi zasadami inżynierii tkankowej (zob. ramkę na s. 177). Hipotetycznie, można by zdążyć do formy konkretnego mięśnia, uzyskując poledwice, szynkę i boczek, jednak póki co problem ten pozostaje nierozwiązany nawet na gruncie medycyny, gdzie doskonale naśladowanie struktury organów jest kwestią życia i śmierci, a nie kulinariów (zob. rozdział IV). Pierwszym krokiem na drodze ku mięsu *in vitro* stał się więc burger z mięsa mielonego, które ze zrozumiałych względów łatwiej naśladować: wystarczy wyprodukować „ogólny mięsień” i go przemielić. Prototypy znane są od 2013 roku, a w grudniu 2020 roku pierwsza restauracja – „1880” w Singapurze – wprowadziła burgery z mięsem *in vitro* do stałej sprzedaży^[21*].

Trchę wbrew nazwie, rolnictwo „komórkowe” dotyczy też jednak produktów nieskładających się z komórek: jak mleko albo jaja^[22*]. W tych przypadkach procedura jest zupełnie inna: celem jest „po prostu” odtworzenie danego produktu, składnik po składniku. Przykładowo, „**mleko *in vitro***” zawiera kazeinę, czyli głównie białko mleka krowiego, produkowaną w komórkach bakterii lub drożdży genetycznie zmodyfikowanych tak, aby wytwarzały „krowią kazeinę”: aminokwas po aminokwasie. Przez pierwsze lata tłuszcze mleka krowiego zastępowane były przez oleje roślinne, ale tendencja jest taka, by stopniowo eliminować kompromisy: pod koniec 2019 roku Perfect Day Foods^[23*], jedna z firm produkujących laboratoryjne mleko, ogłosiła pierwsze eksperymenty z hodowlą krowich komórek tłuszczowych (adipocytów), aby pobierać z nich mieszaninę „naturalnych” tłuszczów zwierzęcych. Nawiasem mówiąc, ponieważ mleko takie jest projektowane „od zera”, można było od początku zrezygnować z kłopotliwej dla wielu osób laktozy, i nie znalazłem żadnej firmy rozwijającej mleko *in vitro*, która by dodawała laktozę do swoich prototypów –

choć byłoby to ukłonem w stronę realizmu. Ostatecznie plan jest taki, aby dążyć do stuprocentowej zgodności składu mleka *in vitro* z prawdziwym mlekiem krowim – co już teraz oznacza choćby możliwość wytwarzania na jego bazie nabiału: istnieją sery robione z mleka *in vitro*.

Pojawia się tu pytanie zaiste filozoficzne: czy produkt, który ma dokładnie taki sam skład, co mleko, ale powstał na drodze manipulacji laboratoryjnych, naprawdę jest mlekiem? Nie mówimy tu wszak o sojowym czy migdałowym substytucie, w odniesieniu do których od 2017 nie wolno w Unii Europejskiej określać terminem „mleko”. Dużo lepszym porównaniem byłby syntetyczny diament (będący pod każdym względem fizykochemicznym „po prostu diamentem”) albo „aromat identyczny z naturalnym” (który jest, o ile nie doszło do oszustwa, dokładnie tą samą cząsteczką chemiczną, co aromat „naturalny”). Ostatecznie zadecyduje pewnie głos konsumentów.

Ruch miłośników „bezzwierzęcych produktów odzwierzęcych” trafił na podatny grunt. Z każdym rokiem rośnie nasza świadomość tego, w jakich, tak naprawdę, warunkach powstają jaja, mleko czy mięso, zaś nie wszyscy chcą albo mogą zostać weganami. Niektórzy co odważniejsi analitycy przewidują, że hodowla zwierząt może w przyszłości całkowicie zniknąć, a badania opinii publicznej nieodmiennie pokazują, że gdyby mięso było nie do odróżnienia od „naturalnego” i nieprzesadnie kosztowne, znaczna większość mięsożerców byłaby zainteresowana przestawieniem się na nie.

Co ciekawe, rolnictwo komórkowe bardzo „się lubi” z ruchem „wytwórców”. Wiele początkujących firm, jak choćby wspomniana wyżej Perfect Day Foods, powstało w warunkach garażowych, a ich twórcy często sympatyzują z ideami wolnego dostępu do wiedzy i „otwartej nauki”. Mało tego, nietrudno znaleźć też powiązania z nastrojami antysystemowymi (stąd na przykład popularność słowa „haker” i „hakowanie” – samą ideę rolnictwa komórkowego określa się czasem jako „zhakowanie łańcucha produkcji spożywczej”^[48]) i, nie bójmy się tego powiedzieć, socjalistycznymi. We wstępie do tego rozdziału nie przez przypadek nawiązałem do Marksa – wielu przedstawicielom ruchu wytwórców chodzi nie tylko o przemianę technologiczną, ale i społeczną. Popularna jest więc wizja, że historia zatoczy ciekawe koło: od indywidualnej hodowli zwierząt, przez hodowlę przemysłową, następnie przemysłowe rolnictwo komórkowe, aż po jego wersję indywidualną. W przyszłości każdy miałby zostać „mięsnym hakerem” albo „nerd-farmerem” (to pierwsze wymyśliłem, to drugie widuję tu i ówdzie), hodując swoje kotlety i zboża w garażowych inkubatorkach.

Nie gdzie indziej, jak na samym MIT powstał Personal Food Computer – „osobisty komputer spożywczy”^[49]. Autorzy opisują go jako „mieszczącą się na stole platformę technologiczną do rolnictwa w warunkach kontroli środowiskowej, która dzięki systemom robotycznym kontroluje i monitoruje klimat, energię i wzrost roślin wewnątrz specjalnej komory hodowlanej”. Projekt jest „otwarty”: w pełni udostępnione są wszystkie schematy konstrukcyjne, listy części i przepisy dla drukarek 3D.

Jak daleko można się posunąć z tego typu myśleniem? Niektórzy uważają, że dowolnie daleko. W 1854 roku Henry David Thoreau opublikował zbiór esejów zatytułowany *Walden, czyli życie w lesie*, w którym zawarł swoje refleksje po dwóch latach samotniczego życia w samodzielnie zbudowanej chatce nad jeziorem. Książka ta stała się potężnym źródłem inspiracji dla głosicieli „ucieczki od cywilizacji”, „prostego życia” i, zwłaszcza w ostatnich dekadach, „zielonego anarchizmu”. Problem z „ucieczką od cywilizacji” polega na tym, że gdyby chciał ją przeprowadzić konsekwentnie, trzeba by zrezygnować z wszystkiego, co ta cywilizacja oferuje. Trudno poważnie potraktować człowieka „obrażonego na korporacje”, który tweetuje na ten temat ze swojego iPhone’a, popijając brazylijską kawę dogodnie przykrytą plastikowym wieczkiem, aby przypadkiem nie nakapała na jego uszyte w Bangladeszu gatki. Podobne kontrowersje budzą rozmaite, rozsiane po świecie „ekowioski”, które w swoich własnych granicach zachowują piękno i harmonię, jednak zbudowane są z materiałów wyprodukowanych we wszystkich zakątkach świata, często z rażąco negatywnym wpływem na środowisko.

Jedynym naprawdę konsekwentnym kierunkiem realizacji tej wizji jest anarchoprymitywizm: ruch mający na celu autentyczny powrót do „stanu dzikiego”, również na poziomie psychologicznym. W ostatnich latach sporą popularnością cieszą się kanały na YouTube prezentujące postępy w rozwijaniu „prymitywnej technologii”. Ich ojcem chrzestnym jest Australijczyk John Plant, założyciel kanału Primitive Technology, który od 2015 roku buduje własnymi rękami coraz bardziej zaawansowane domy, piece, naczynia ceramiczne sieci do połowu i materiały tkane, rozpoczynwszy od narzędzi kamiennych^[24*].

To inspirujący kierunek, ale – trzeba przyznać – nie dla każdego. Nie brakuje więc ludzi, którzy próbują na własną rękę znaleźć jakiś sensowny kompromis pomiędzy pełnym uczestnictwem w globalnym przepływie towarów i usług oraz odłączeniem od tego, co uważają za

moralnie i estetycznie nie do przyjęcia w cywilizacji technicznej i kapitalizmie. W końcu uczestniczyć w cywilizacji też można mniej lub bardziej mądrze.

Na przestrzeni ostatniego stulecia powstawały więc niezliczone komuny, mniej czy bardziej zainspirowane romantycznym eksperymentem Thoreau, akceptujące jednak przynajmniej część współczesnego dorobku cywilizacyjnego. Trudno bowiem liczyć na to, że da się zupełnie niezależnie od globalnego rynku wytopić stal i wyprodukować z niej blachę falistą, wydrukować obwód elektroniczny albo zsyntetyzować insulinę dla anarchistów-diabetyków. Hm... A może jednak? W ostatnich latach stopniowo zaczął wyłaniać się plan kontrolowanego odłączenia od cywilizacji: po „podkradnięciu” z niej technologii pozwalających na osiągnięcie „samodzielności na wysokim poziomie” miałyby nastąpić stopniowe zerwanie więzi z Systemem.

W 2008 roku pochodzący z Polski fizyk jądrowy Marcin Jakubowski wygłosił na Uniwersytecie Columbia w Missouri wykład, na którym opisał ideę Global Village Construction Set (GVCS, „Globalny Zestaw do Konstrukcji Wioski”). Jest to zestaw instrukcji pozwalający na samodzielne skonstruowanie 40 maszyn (później listę rozszerzono do 50) oraz wykonanie przy ich pomocy w pełni samowystarczальной wioski. Pierwszą maszyną na liście jest drukarka 3D, pomagająca w konstrukcji następnych urządzeń: plan jest taki, aby po początkowej inwestycji w wysokości około 10 tysięcy dolarów możliwe było stopniowe rozbudowywanie bazy technologicznej, wyłącznie w oparciu o już istniejące urządzenia oraz dostępne na otwartej licencji schematy konstrukcyjne. Lista 50 urządzeń składających się na GVCS robi wrażenie; znajdują się na niej między innymi: turbina wiatrowa, traktor, wycinarka laserowa, piec chlebowy, akumulator, rozdrabniacz do słomy, betoniarka, silnik parowy, piec indukcyjny, spawarka i wiele innych. „Zestaw” ma docelowo pozwolić na zbudowanie „małej, samowystarczальной cywilizacji ze współczesnymi wygodami”.

W kolejnych latach Jakubowski zaczął wprowadzać swój plan w życie na farmie-fabryce o nazwie Factor e Farm, położonej w odludnej okolicy na terenie stanu Missouri. Pierwsza maszyna, prasa do tworzenia „cegieł ziemnych” (technicznie: „sprasowanych bloków ziemnych”, CEB), powstała już w 2007 roku. W 2013 roku zbudowano pierwszy dom, posługując się wyłącznie urządzeniami wytworzonymi na miejscu. W 2018 roku, kiedy powstał ostatni syntetyczny raport prezentujący stan prac^[50], stopień rozwoju

„Zestawu” oszacowano na 33%. Na liście całkowicie opracowanych znajdowały się tylko cztery urządzenia: drukarka 3D, prasa do CEB, „Microhouse” (dom z cegieł CEB) oraz... rozgałęźnik elektryczny („złodziejka”). Blisko ukończenia są między innymi: rozdrabniarka do gleby, traktor i frezarka CNC.

Trudno zaprzeczyć, że tego typu minicywilizacja, wytwarzająca swoje własne zaplecze technologiczne, to wizja szczególnie atrakcyjna dziś, gdy z jednej strony coraz częściej słyszymy o konieczności „odłączenia się” od technologiczno-informacyjnej Bestii, a z drugiej wcale nie mamy ochoty rezygnować z komfortów, którymi nas ona wabi. Trudno też zaprzeczyć, że detale tej wizji pozostają niejasne – co z elektroniką i internetem, co z medycyną, co z ropą naftową i metalami rzadkimi? Dalej: czy aby odpowiednio duża „wioska open source” nie zacznie po cichu odtwarzać wszystkich błędów cywilizacji? Jak długo trzeba by czekać na pojawienie się w niej, zupełnie spontanicznie, zawodu bankiera?

Osoby pokroju Jakubowskiego – przynajmniej ci, którzy są na tyle trzeźwi, że mogą pokazać jakieś udane realizacje – nie są oczywiście ślepi na takie zarzuty. Celem samym w sobie są wszystkie kroki pośrednie, jak choćby instrukcje pozwalające każdemu człowiekowi z dostępem do internetu na skonstruowanie prasy do cegieł ziemnych albo frezarki CNC za ułamek ceny rynkowej. Opisane w tym rozdziale urządzenia pozwalające na samodzielne wytwarzanie czegoś, co dotychczas powstawało wyłącznie w wielkich fabrykach, dla mieszkańca Kalifornii mogą być źródłem ciekawej rozrywki, ale dla milionów ludzi na świecie jedyną szansą na dostęp do niektórych dobrodziejstw cywilizacji.

II.2. Świat treści tylko dla ciebie

Po trudach „kontrolowania materii”, opisanych w poprzednim rozdziale, mogłoby się wydawać, że w erze komputerów kontrolowanie informacji będzie bułką z masłem. Cóż to bowiem za problem wygenerować parę zdań albo obraz? Przecież to nic nie kosztuje i dokonuje się w mgnieniu oka! O, proszę: „kotlet”. Okazuje się jednak, że napisanie przekonującego opowiadania bez udziału człowieka jest znacznie trudniejsze niż stworzenie przekonującego burgera wołowego bez udziału krowy.

Generowanie sensownej treści to zadanie lustrzane wobec zrozumienia treści – czyli wydobycia z niej „sensu”. Rzecz w tym, że

nikt tak naprawdę do końca nie wie, jak to się dzieje, że człowiek patrzący na pokój „rozumie” otaczającą go rzeczywistość; albo czym właściwie jest „rozumienie” tekstu albo obrazu i czym jest ich „sens”. Wiemy, że komputery świetnie operują symbolami i potrafią dokonywać cudów z ciągami cyfr i znaków albo listą komend napisanych w ustalonym języku programowania. Są jednak kompletnie bezradne wobec zdania „Mój syn poszedł sam do przedszkola, a zaczęło mocno padać” – w tym sensie, że człowiekowi czytającemu te słowa natychmiast „uruchamia się” całe morze skojarzeń, pytań, emocji i hipotetycznych scenariuszy, podczas gdy dla „naiwnego” komputera, nawet posiadającego do dyspozycji bazę danych wszystkich słów języka polskiego, nie ma istotnej różnicy między zdaniem powyższym a tym: „Mój kwadrat poszedł sam do muzeum, a zaczęło zielono opodatkować”.

Można by napisać całe książki o naszych próbach „nauczenia komputerów świata”: wszczepienia im czegoś w stylu zdrowego rozsądku, mapy elementarnych skojarzeń, reprezentacji sensu słów, wiedzy o rzeczywistości ludzkiej i tak dalej. Nawet samo nazwanie problemu stanowi problem sam w sobie. W największym uproszczeniu można by powiedzieć, że spierają się tu dwa podejścia, które można by określić jako „całościowe” i „na sztuczki”.

Podejście „całościowe” miałyby polegać na „wyhodowaniu” w krzemie samodzielnej, być może „świadomej” (cokolwiek to znaczy) istoty, oddziałującej ze światem, która dopiero wtedy nauczyłaby się czytać i pisać. Idea jest więc taka, że sensowny tekst może być generowany wyłącznie przez istotę, która „doszła do sensu” taką samą drogą, co człowiek: poprzez długotrwały rozwój kompleksu ciało–umysł. Niektóre z tych wątków pojawiły się już przy okazji omawiania architektur kognitywnych (zob. ramkę na s. 49) i nie będę ich teraz powtarzał. Ten temat odkładam na bok – nie dlatego, że nie wierzę w jego realizację, a raczej dlatego, że przy jego omawianiu miałbym do pokazania niewiele konkretnych i niewiele sukcesów.

Metoda „na sztuczki” też już gościła na kartach tej książki. Wyposażanie robotów w „sztuczną inteligencję”, o czym pisałem w rozdziale I, dokonuje się wszak właśnie przez dodawanie im kolejnych funkcji, aż imitacja inteligencji stanie się wystarczająco dobra. W świecie rozumienia i generowania języka naturalnego, jak łatwo się domyślić, metoda ta cieszy się wielkim powodzeniem. Cóż, istnieją już programy komputerowe *naprawdę dobrze* tłumaczące pomiędzy niektórymi językami, na przykład między angielskim i hiszpańskim. Istnieją też generatory języka naturalnego, tworzące

już w pełni funkcjonalne proste teksty, na przykład małe formy dziennikarskie, ale też swobodniejszą prozę na zadany temat. O tym pogadamy nieco więcej.

Aby zrozumieć, co dzieje się w „mózgu” tego typu programu, trzeba by zaprzyjaźnić się z różnymi sztuczkami stosowanymi dziś przez programistów. Jest tego... mnóstwo! „Sztuczna inteligencja” to, jak już wspominałem, tak naprawdę gigantyczny zbiór rozmaitych drobnych narzędzi, metod i pomysłów, z których korzysta się w zależności od potrzeb. Zgodnie z duchem tej książki, chciałbym pokazać wam tylko kilka przykładów, które nie złożą się raczej w wyczerpujący wykład tego, czym jest dzisiaj „AI”, ale które – mam nadzieję – pokażą przynajmniej, jakiego typu „zębatki” będą siedzieć w „mózgach” maszyn generujących wam w przyszłości opowieści, filmy i gry komputerowe.

II.2.1. Słowa, słowa, słowa...

Od kiedy tylko dowiedziałem się o wektoryzacji słów^[25*], stwierdziłem, że muszę o niej opowiedzieć w tej książce. Idea ta na pierwszy rzut oka nie wydaje się może szczególnie spektakularna, ale stanowi doskonałą ilustrację tego, w jaki sposób rozmaite drobne techniki po trochu przelewają w świat zer i jedynek bogactwo świata ludzkiego. Co więcej, leży ona u podstaw wielu innych, bardziej zaawansowanych technologii, których opisu nie zmieścilibyśmy już w tej książce.

To dłuższa historia, ale warto.



Słowa jako wektory

Zacznijmy od przypomnienia, jak „stare, dobre algorytmy” reprezentowały słowa. Była to po prostu długa lista, na której każde słowo stanowiło osobny obiekt, niezwiązany w żaden sposób z pozostałymi. O, tak:

000001:	
000002:	abakus
000003:	abażur
000004	abecadło
...	

Jest to więc wielki „worek”, w którym leżą, bez żadnego związku ze sobą nawzajem, miliony słów. Zasadnicza idea „wektorów słów” jest taka, aby słowa rozmieścić w jakiejś wymyślonej przestrzeni tak, aby ich położenie dostarczało nam użytecznej wiedzy. Prosta analogia – wyobraźmy sobie alfabetyczną listę miast: Aachen, Abadan, Aberdeen... Sąsiedztwo na takiej liście jest czymś zupełnie przypadkowym. A co, gdybyśmy posłużyli się układem geograficznym? Do każdego miasta dodajmy informację o jego położeniu na globie, co można zrobić przy pomocy dwóch liczb: długości i szerokości geograficznej. O, tak:

Aachen:	6,08	50,76
Abadan:	48,30	30,33
Aberdeen:	-2,18	57,25
...		

Wbrew pozorom ten prosty zabieg dostarczył nam sporo informacji. Przypuśćmy, przykładowo, że byliśmy już kiedyś w Szkocji i spędziliśmy majówkę w Aberdeen – gdzie, dodajmy, średnia temperatura w maju to 9°C. Choć miasto nas zachwyciło, nieprędko zapomnimy również klimatyczną szkocką mżawkę. Tymczasem, przy jakiejś okazji, znajoma informuje nas, że spędziła swoją majówkę w mieście o nazwie Aboyne. Słowo to nic nam nie mówi. Ponieważ jednak zawsze nosimy przy sobie tablice geograficzne, szybko odnajdujemy w niej „namiary” na Aboyne: [-2,13; 57,13]. A ponieważ jesteśmy też komputerami, bez trudu obliczamy odległość Aboyne od wszystkich innych miast, po czym okazuje się, że to przecież rzut kamieniem od Aberdeen! Dzięki tej prostej informacji jesteśmy w stanie błyskawicznie zareagować i zapytać koleżankę, czy miała ze sobą ciepły sweter i płaszcz przeciwdeszczowy oraz jak jej smakowały szkockie piwa. Prosta wiedza o odległości między miastami pozwoliła nam dokonywać *przewidywań* – bliskie sobie miasta przynajmniej pod pewnym względami raczej będą do siebie bardziej podobne niż miasta odleglejsze.

Pociągnijmy tę metaforę odrobinę dalej. Przypuśćmy, że do naszej „tablicy miast” dodajemy kolejną kolumnę, reprezentującą tym razem liczbę mieszkańców.

Aachen:	6,08	50,7	248960
		6	
Abadan:	48,3	30,3	231476
	0	3	
Aberdeen:	-2,1	57,2	200680
	8	5	
Aboyne:	-2,1	57,1	2910
	3	3	

Póki co rozmieszczaliśmy nasze miasta w przestrzeni dwuwymiarowej; możemy wyobrazić sobie, że jest to mapa świata rozłożona na podłodze. Teraz dodaliśmy trzeci

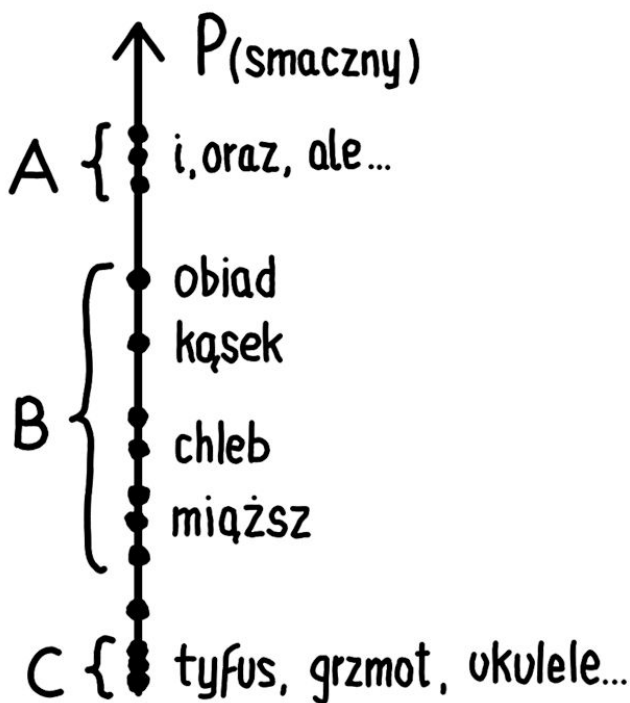
wymiar: nasze miasta unoszą się więc w różnej odległości od podłogi: wielkie miasta pod sufitem, a małe wioski tuż przy podłodze. To kolejne źródło informacji, a rozmaite „chmury”, które wyłoniły się dzięki dodaniu kolejnego wymiaru, stają się coraz bardziej pomocne: w mojej nowej trójwymiarowej przestrzeni pojawiła się na przykład szczególnego rodzaju „bliskość” pomiędzy trzema pierwszymi miastami na powyższej liście: na tyle dużymi, że można się już w nich spodziewać przynajmniej kilku interesujących budynków o niebanalnym przeznaczeniu, a być może i ciekawej historii, jednak wciąż na tyle małymi, że nie są to tętniące życiem metropolie, których centrum nigdy nie zasypia.

Nie jest to oczywiście metoda bezbłędna i ścisła matematycznie: miasta mają swój charakter i swoją historię, których nie da się zawrzeć w kilku liczbach. Idea jest jednak taka, że każdy kolejny parametr – rozmieszczający nasze miasta w coraz to wyżej wymiarowych przestrzeniach – zwiększa prawdopodobieństwo, że odległość między miastami coś nam o nich mówi. Dodajmy teraz, że dla matematyka słowo „wektor” oznacza po prostu uporządkowaną listę liczb, jak te wymienione wyżej, a wyobrażanie sobie przestrzeni, powiedzmy, 100-wymiarowej, wcale nie jest konieczne, aby się nimi swobodnie posługiwać. W zupełności wystarczy intuicja trójwymiarowa.

Co jednak ze słowami? I tu kryje się prawdziwe piękno „wektoryzacji słów”: idea jest taka, żeby słowa znalazły się blisko wtedy, jeśli – uwaga! – są używane podobnie. Inaczej mówiąc, „kontekst to wszystko”. Zasadnicza metoda polega na przeczesaniu każdego gigantycznego zbioru tekstowego^[26*] i sprawdzaniu, jakie jest otoczenie każdego z występujących w nim słów. I... tyle. Najbardziej znany algorytm tego typu nosi nazwę Word2Vec i został opisany^[51] w 2013 roku przez grupę programistów Google pod kierunkiem Tomasa Mikolova.

Zacznijmy od prostego przykładu. Przypuśćmy, że interesuje mnie, jak używane jest przez Polaków słowo „smaczny”. Zaczynam od zgromadzenia olbrzymiego zbioru polskich tekstów – szczęśliwie istnieje już stworzony przez polskich językoznawców Narodowy Korpus Języka Polskiego o objętości ponad 1,5 miliarda słów, składający się z tekstów książek, artykułów prasowych, stron internetowych i nagrań rozmów^[27*]. Wyszukuję więc w nim wszystkie wystąpienia słowa „smaczny” (oraz jego odmian typu „smaczna” i „smaczny”), odnotowując, jak często wszystkie pozostałe słowa występują po jego lewej i prawej stronie. Rekordzistami, nawiasem mówiąc, są „jest” (i jego odmiany) po lewej stronie oraz „i” po prawej. W kategorii rzeczowników, występujących niemal zawsze po prawej stronie, czołówka to: „obiad”, „kąsek”, „posiłek” i „poczęstunek”. Nieco niżej: „chleb”, „owoc”, „napój”. I tak dalej.

Co istotne, badanie to dostarczyło mi nie tylko informacji o słowie „smaczny”, ale również o *każdym innym słowie języka polskiego*. Mogę teraz przyporządkować każdemu z nich liczbę, odpowiadającą liczbie przypadków, kiedy wystąpiło ono po prawej stronie słowa „smaczny”; oznaczmy ją może $P_{(\text{smaczny})}$. To pierwszy wymiar mojej przestrzeni wektorowej:

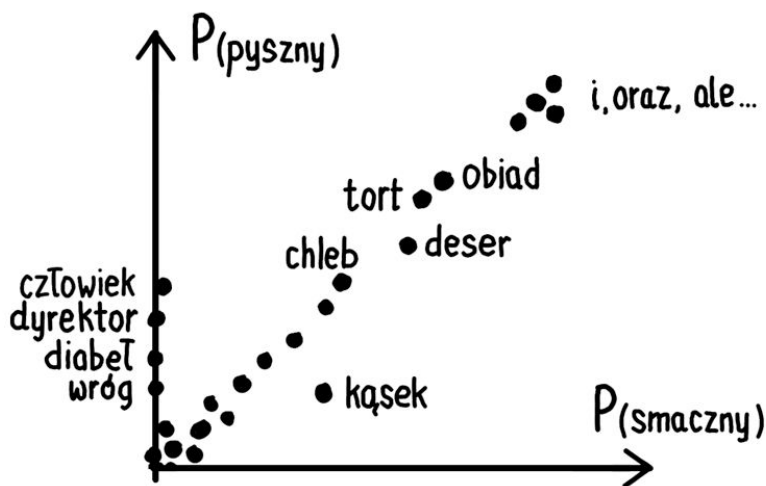


Wyraźnie widać tu trzy główne grupy. Dwie są niezbyt interesujące: pierwsza z nich to te słowa, które „lubią się z wszystkimi słowami” (A) i przez to nie dostarczają zbyt cennych informacji. Druga to te, które nigdy lub prawie nigdy nie wystąpią w otoczeniu słowa, które badamy (C): również i te nie są zbyt interesujące. Najciekawsze rzeczy dzieją się pośrodku skali (B): to te słowa, które występują względnie często, są jednak rzadsze od „bestsellerów” danego języka, jak popularne spójniki z grupy A.

Idźmy dalej, biorąc na warsztat słowo „pyszny”. Przeprowadzamy analogiczne badanie i uzyskujemy dość podobną na pozór mapę:



Choć mogłoby się wydawać, że „smaczny” i „pyszny” to synonimy, szybko okazuje się, że są pewne słowa znajdujące się dość „wysoko” na wykresie $P_{(\text{pyszny})}$, a dość nisko na wykresie $P_{(\text{smaczny})}$. Ponieważ zaś idea jest taka, że próbujemy myśleć o tych parametrach jako o wymiarach pewnej przestrzeni, nałożmy jeden na drugi, tworząc zwykłą, swojską przestrzeń dwuwymiarową:



No i proszę. Zupełnie spontanicznie wyłoniła się grupka rzeczowników, które „mogą być pyszne”, ale „nie mogą być smaczne”; należą do nich „człowiek”, „dyrektor”, „diabeł”, „wróg” i wiele innych. No, oczywiście! Przecież słowo „pyszny” może oznaczać również: „przepełniony pychą”! Zauważmy, że ta subtelność, którą w modelu „worka słów” trzeba by zakodować ręcznie, tutaj wyłoniła się zupełnie spontanicznie. Zauważmy również przy okazji, że rozdzielenie „tego samego słowa” (tej samej sekwencji liter) na różne znaczenia wcale nie jest oczywiste. Książkowy przykład to „zamek”, który zwykle przedstawia się jako posiadający „dwa różne znaczenia”: budynek warowny albo zamknięcie, na przykład drzwi. *Uniwersalny słownik języka polskiego* PWN-u podaje jednak aż pięć głównych znaczeń tego słowa: „warowna budowla”, „urządzenie do zamykania drzwi, szuflad, walizek”, „zapięcie zrobione z zaczepiających się o siebie ząbków”, „urządzenie do łączenia lub zabezpieczania elementów maszyny, np. zamek zaworowy” oraz „mechanizm broni palnej służący do zamykania na czas wystrzału tylnej części lufy”. W starych, dobrych czasach twórcy programów tłumaczeniowych musieli żmudnie rozdzielać każde słowo na jego poszczególne znaczenia, których jest... nie wiadomo właściwie ile.

Metoda wektorów słów załatwia wszystkie te problemy za jednym zamachem, a każdy dodatkowy wymiar tylko pogłębia subtelność uzyskanego grupowania. Zauważmy, że póki co udało nam się względnie czysto wydzielić spośród wszystkich słów:

- a) „te, które mogą być smaczne i pyszne”,
- b) „te, które mogą być pyszne, ale nie mogą być smaczne”,
- c) „te, które nie mogą być ani takie, ani takie”.

To jednak dopiero dwa wymiary! Idźmy dalej. Po dodaniu wymiaru $P_{\text{(słodki)}}$ okazałoby się, przykładowo, że dotychczasowa jednolita środkowa smuga, reprezentująca smaczne potrawy, zaczyna rozsuwać się na podgrupki potraw „zwykle słodkich” i „zwykle niesłodkich”. Dużo częściej występuje wszak para słów „słodki deser” niż „słodki bigos”! Co więcej, podgrupka „przepełnionych pychą istot człekokształtnych” również zaczęłaby się rozsuwać, ponieważ (sprawdziłem to) częściej mówimy „słodki” o „człowieku” albo „dyrektorze” niż o „wrogu” i „diabie”.

Tego typu subtelne efekty – dzięki którym analiza każdego słowa wprowadza kolejne subtelne rozróżnienia do naszej rosnącej „przestrzeni słów” – są właśnie odpowiedzialne za potęgę „wektoryzacji słów”. Nie musimy żmudnie uczyć komputera, co to znaczy „dyrektor”. „Sens” (choć ostrożnie z tym pojęciem) tego słowa, jego odrębność, jego „sąsiedztwo” i typowe sposoby używania wyłaniają się z milionów przypadków użycia przez zwykłych ludzi (dlatego tak ważne jest, aby baza danych zawierała teksty pisane normalnym językiem przez przeciętnych ludzi).

Jak długo można przeprowadzać tego typu procedurę? Teoretycznie, dowolnie długo – najbardziej oczywistym rozwiązaniem byłoby „przeczesanie” bazy tyle razy, ile słów występuje w słowniku danego języka. W takim przypadku każde słowo zostałoby „krzyżowo” ocenione ze względu na „pokrewieństwo” z każdym innym słowem; i to przynajmniej dwukrotnie, ponieważ za każdym razem sprawdzalibyśmy sąsiada lewego i prawego (w rzeczywistości zaś nawet jeszcze więcej, ponieważ okazuje się, że opłaca się badać nawet „okienko” o rozpiętości pięciu słów w każdą stronę).

Szybko pojawi się jednak poważny problem obliczeniowy. Jedna ze stosowanych przez Googla baza danych „wektorów słów” (Google News Word2Vec) zawiera około 3 milionów różnych słów. Gdybyśmy chcieli wygenerować na jej podstawie „przestrzeń słów”, miałaby ona 6 milionów wymiarów (dla najprostszego przypadku, kiedy dla każdego słowa sprawdzamy jego najbliższych sąsiadów z lewej i z prawej). To nie jest wygodne i nie pozwala na szybki dostęp do danych – nawet jeżeli ma się do dyspozycji zasoby obliczeniowe i dyskowe Googla. Tu wkraczają matematycy.

Nie będziemy bliżej przyglądać się stosowanej przez nich magii, ale dość powiedzieć, że „redukcja wymiarowa” to znany problem w matematyce i istnieje cała biblioteka rozwiązań pozwalających na „upakowanie” przestrzeni pierwotnie, powiedzmy, 100-wymiarowej do postaci przestrzeni 20-wymiarowej. Cała sztuczka polega na tym, aby w procesie tym zgubić jak najmniej informacji. Na szczęście język jest bardzo „powtarzalny” i kiedy już sprawdzimy, jak wszystkie słowa „mają się” do słów, powiedzmy, „kaszałat” i „finwał”, to kolejna tura badania, skupiona na słowie „pletwał”, nie doprowadzi do wielkiej rewolucji. Dobrze zaprogramowany program komputerowy będzie „wiedział”, czy informacja dostarczona w danym kroku wymaga powołania do życia kolejnego wymiaru, czy też wystarczy nieznaczne „przemeblowanie” wektorów w dotychczasowej przestrzeni.

Po latach doświadczeń okazało się, że rozsądnym kompromisem jest przestrzeń o około 300–400 wymiarach. Jeżeli więc zastanawiacie się, jak współczesne programy „rozumiejące” język naturalny, na przykład programy tłumaczące i generujące tekst, „wyobrażają” sobie słowa, odpowiedź brzmi: jeżeli korzystają z algorytmów typu Word2Vec, to jako uporządkowaną listę około 300 liczb.

Uff. Mamy to.

Co z tego wszystkiego wynika? Zaskakująco dużo.

Niektóre rezultaty możliwe do uzyskania dzięki wektoryzacji słów wyglądają, słowo daję, magicznie. Jednym z nich jest fakt, że przestrzenie wygenerowane w różnych językach będą miały bardzo podobną strukturę, dzięki czemu wektory słów mogą być – i są – wykorzystywane przy tłumaczeniach. Okazuje się, że słowa znajdujące się „w podobnych miejscach” w językach A i B bardzo często są swoimi

słownikowymi tłumaczeniami. „Kot” ma się do „pies” w przestrzeni trenowanej na polskim korpusie, praktycznie dokładnie tak samo, jak „cat” do „dog” w przestrzeni trenowanej na korpusie angielskim. Gdy zaś poszerzyć nieco horyzonty i badać nie pojedyncze słowa, a frazy, to okaże się, że fraza „Good morning” ma się do „Goodbye” tak, jak „Dzień dobry” do „Do widzenia”.

Inny błyskawiczny rezultat: popularne literówki nie nastrożają żadnego problemu: słowa „Schwarzenegger”, „Schwarzeneger” i „Shwarzenegger” stosowane są dokładnie tak samo i lądują w tym samym miejscu przestrzeni. Jeżeli zastanawialiście się kiedyś, jakim cudem dzisiejsze aplikacje tak łatwo radzą sobie z literówkami, to już wiecie. To nie długie listy popularnych literówek, pracowicie wklepywane przez stażystów (choć listę taką ma od pradawnych czasów choćby edytor tekstu Microsoft Word) – to magia wektorów słów.

Jeszcze ciekawsze są „miękkie” konsekwencje „wektoryzowania” słów. Zwróćmy uwagę, że każdy „przełot” algorytmu przez następne słowo wzbogaca bazę danych o kolejny aspekt doświadczenia ludzkiego, zaklętego w języku. Uczony w ten sposób algorytm „wchłania” więc po trochu elementy wiedzy powszechnej (i specjalistycznej), które w „starych, dobrych” systemach trzeba by w nieskończoność wklepywać ręcznie. Od 1984 roku rozwijany jest projekt o intrygującej dla polskiego ucha nazwie Cyc^[28*], który miał pierwotnie stanowić „bazę danych wiedzy powszechnej”: zbiór zdań typu „Woda jest mokra” albo „Drzewa są roślinami”. W 2017 roku baza Cyc zawierała 25 milionów takich „cegiełek”, ręcznie wklepanych przez kolejne pokolenia entuzjastów. Procedury podobne do wektoryzacji słów w pewnym sensie „wchłaniają” tego typu elementarne prawdy o świecie zupełnie automatycznie.

Prosty przykład: miasta bliskie sobie geograficznie najprawdopodobniej znajdują się blisko siebie również i w naszej 300-wymiarowej przestrzeni wektorowej, po prostu dlatego, że ludzie dużo częściej wymieniają jednym tchem Kraków i Wieliczkę niż Kraków i Aberdeen. Blisko siebie znajdują się też imiona królów, którzy panowali w podobnym czasie, nazwiska sportowców, którzy latami grali w jednej drużynie, albo... nazwy potraw o podobnym składzie i smaku.

Powiązania te mogą być jeszcze bardziej subtelne. Ot, liczba 1939 dla „naiwnego” algorytmu nie różni się niczym od liczby „2939” – no może poza czysto numeryczną różnicą wynoszącą 1000. W przestrzeni

wektorowej wytrenowanej na dużym zbiorze tekstów liczba ta jest jednak „znacząca” – jest wszak „osadzona” znacznie bliżej słów „wojna” czy „wybuch” niż liczba „2939”. Odpowiednikiem tej „bliskości” na poziomie świadomości człowieka są chyba skojarzenia. W przestrzeni wektorowej słów istnieją więc „lata kojarzące się z wojną”, „nazwiska kojarzące się groźnie” albo „miasta kojarzące się ze słońcem”.

Aż kusi, żeby wiedzę tę jakoś wykorzystać.



Mądrość miliona tekstów

Wyobraźmy sobie następujące „zadanie”: „1914 ma się do 1918, jak 1939 do...?”. Na pytanie to można odpowiedzieć na dwa sposoby: matematycznie albo historycznie. Nie trzeba żadnych szczególnych kompetencji, aby odpowiedzieć na to pytanie „matematycznie”. Dwie pierwsze liczby różnią się o 4, a więc prawidłowa odpowiedź to 1943. Wielu ludzi mogłoby jednak odruchowo odpowiedzieć „1945”. Dotychczas wydawało się, że aby komputer odpowiedział na takie pytanie, trzeba by go żmudnie uczyć nie tylko faktów, ale też osobno wymyślić jakiś sprytny algorytm do wyszukiwania powiązań pomiędzy nimi.

Tymczasem w świecie „słów-wektorów” tego typu zadania „rozwiązują się same”. Co ciekawe, okazuje się, że między parami słów typu „1914”–„1918” oraz „1939”–„1945” wytwarza się pewna szczególna relacja, mająca charakter... geometryczny. Te dwa słowa są od siebie *odległe o pewien określony wektor* – aby więc w przestrzeni wektorowej słów „znaleźć” rok zakończenia jakiejś wojny, należy rozpocząć w miejscu, w którym znajduje się rok jej rozpoczęcia, a następnie „zrobić trzy kroki w lewo i w górę” (przy czym instrukcja ta jest w rzeczywistości 300-wymiarowa). To nie żart, ta metoda naprawdę działa!

Podobne zależności przestrzenne łączą na przykład państwa z nazwami ich stolic albo wynalazki z ich wynalazcami. Mówiąc czysto matematycznie: „Warszawa” minus „Polska” równa się „Paryż” minus „Francja” (czyli „strzałka” o początku na słowie „Polska” i końcu na słowie „Warszawa” po „przyłożeniu” jej początku do słowa „Francja” wskazałaby swoim końcem na słowo „Paryż”). Automatyczna umiejętność dopowiadania stolic świata to jednak nie wszystko. Algorytmy wektoryzujące słowa wychwytyją również znacznie bardziej subtelne skojarzenia i bywają źródłem nieustającej zabawy dla programistów, którzy prostą metodą dodawania wektorów sondują, jakie skojarzenia buzują pod powierzchnią naszej zbiorowej świadomości.

Oto parę rzeczywistych wyników, uzyskanych poprzez „odpytywanie” 300-wymiarowej przestrzeni wektorowej słów, „nauczonej” na podstawie gigantycznej bazy krótkich tekstów dziennikarskich dostępnych w serwisie Google News. W nawiasie kwadratowym wymienione są najlepsze, w „opinii” algorytmu, uzupełnienia^[52].

dom→dach	zamek →	[kopuła, dzwonnica, wieża]
kolano→noga	łokieć→	[ręka, ramię]

Zdarzają się też ciekawe anomalie.

król→królowa mężczyzna→ [kobieta, próba porwania, nastolatka, dziewczyna]

Niezwykła jest jednak zdolność tego typu algorytmów do „bezwysiłkowego” poruszania się choćby w świecie ludzkich emocji:

miłość→obojętność	strach	[apatia, bezduszość, nieśmiałość, bezsilność,
ć	→	bezczyarność]

„Zrozumienie” subtelności życia ludzkiego to oczywiście bezcenny zasób, kiedy przychodzi do tłumaczenia tekstów albo „sensownego” reagowania na nasze wypowiedzi, jednak nie zapominajmy o potężnym zasobie czysto merytorycznej wiedzy zakłętej w przestrzeniach wektorowych słów. „Nakierowany” poprzez relację „New York Times→Sulzberger” (Arthur Sulzberger był przez 23 lata kierownikiem The New York Times Company) algorytm wypływa w reakcji na słowo „Fox” cztery nazwiska (Murdoch, Chernin, Bancroft, Ailes), należące do właścicieli lub kierowników Fox News lub News Corp., jej spółki matki. Jeszcze bardziej subtelny jest wynik uzyskany po zapytaniu: „Chiny→Tajwan, Rosja→?”. Trzy pierwsze słowa zaproponowane przez algorytm to „Ukraina”, „Mołdawia” i „Armenia”. Aby człowiek odpowiedział w ten sposób, musiałby nie tylko wiedzieć, że Chiny są dla Tajwanu znacznie potężniejszym, groźnym sąsiadem, silnie ograniczającym jego autonomię polityczną, ale również to, dla jakich państw Rosja pełni podobną rolę. W przestrzeni wektorowej słów tego typu skojarzenia są po prostu „organicznie” zakodowane.

Czyżby więc na nasze kolana trafiła „maszynka” do odkrywania powiązań między faktami? Wielu naukowców uważa, że tak się właśnie stało.

W lipcu 2019 roku w „Nature” ukazał się artykuł^[53], którego autorzy „podsunęli” algorytmowi Word2Vec ponad 3 miliony streszczeń (abstraktów) artykułów naukowych z lat 1922–2018, skupiając się na terminach naukowych mających znaczenie dla nauk o materiałach. Ostatecznie uzyskali 200-wymiarową przestrzeń, w której rozmieścili 500 tysięcy terminów i... siedli do jej „odpytywania”.

Prostym przykładem jest pytanie: jakie materiały znajdują się najbliższe wskazanego związku? Po „rozejrzeniu się” w okolicach nazwy „LiCoO₂” – jest to wzór chemiczny związku stosowanego m.in. przy produkcji elektrod w bateriach litowo-jonowych – okazało się, że algorytm bardzo słusznie rozmieścił w jej pobliżu pięć innych materiałów o podobnych właściwościach, również stosowanych do wytwarzania elektrod. Okazało się też, że algorytm spontanicznie odtworzył układ okresowy, rozmieszczając nazwy pierwiastków w sposób z grubsza odpowiadający ich położeniu na tablicy Mendelejewa.

To jednak dopiero rozgrzewka. Autorzy postanowili sprawdzić, jakie nazwy materiałów trafiły w miejsce odpowiadające materiałom termoelektrycznym, tj. takim, które szczególnie łatwo „pośredniczą” między zjawiskami cieplnymi i elektrycznymi. Obok „starych znajomych” znalazły się tam również związki, które nie były dotychczas wymieniane w literaturze w kontekście termoelektryczności ani żadnego z powiązanych z tym tematem słów kluczowych. Dla dziesiątki szczególnie „obiecujących” materiałów – czyli, powtórzmy, słów, które trafiły w to samo miejsce

tylko dlatego, że były używane w podobny sposób przez pokolenia naukowców – przeszukano istniejące bazy danych pod kątem właściwości fizycznych typowych dla termoelektryków i wygląda na to, że udało się odkryć przynajmniej kilku dobrych kandydatów. Ba, udało się zidentyfikować całe klasy związków, z których żaden nie był dotychczas otwarcie rozważany jako materiał termoelektryczny, a które – jak ostrożnie zauważają autorzy – „dobrze rokuja”. Oczywiście, na tym etapie konieczne byłyby rzeczywiste eksperymenty z próbkami tych materiałów. Zauważmy jednak, jak niezwykle jest to, co nastąpiło: w pewnym sensie suma napisanych przez naukowców tekstów jest „mądrzejsza” od nich samych!

Wróćmy może jednak do konkretnego problemu generowania naturalnego tekstu. Opisane wyżej narzędzie można stosować na dwa podstawowe sposoby: do przewidywania słowa na podstawie kontekstu i przewidywania kontekstu na podstawie słowa. W tym pierwszym przypadku „zagadka” do rozwiązania może brzmieć na przykład: „Nigdy bym cię nie zdradził! Jesteś moim najlepszym [?]”, gdzie [?] oznacza jedno słowo. W tym drugim przypadku zagadka może brzmieć na przykład: „,[??] zamordował [??]” i bywa, że skutki takich eksperymentów bywają interesujące.

Prawda jest jednak taka, że same tylko wektory nie są w stanie wygenerować zbyt interesującego tekstu. Potrzebne jest zupełnie inne narzędzie – coś, co również „nauczy” się języka naturalnego na podstawie jego gigantycznej próby, ale będzie też w stanie reagować na bieżąco na podsuwany mu tekst. W drugiej dekadzie XXI wieku zaczęły powstawać niezliczone odmiany **głębokich sieci neuronowych**, które na poziomie „anatomicznym” nie mają wiele wspólnego z algorytmami typu Word2Vec, jednak realizują tę samą zasadniczą ideę. Na etapie „treningu” algorytm stopniowo uczy się języka, tworząc „dla swoich potrzeb” jego ukrytą reprezentację w kontakcie z gigantyczną bazą tekstu naturalnego. Następnie, na bazie tej reprezentacji może zostać wygenerowany nowy tekst.

Istnieją całe rodziny sieci neuronowych wykonujących tego typu zadania, ukryte pod tajemniczymi nazwami i skrótami – trochę jak w świecie druku 3D (zob. ramkę na s. 81). Najbardziej znanym z nich jest chyba **GPT** (Generative Pre-trained Transformer, czyli „Generatywny, Wstępnie Wytrenowany Transformer”), opisany w 2018 roku przez programistów z założonej przez Elona Muska firmy OpenAI – obecnie jest to jeden z czołowych graczy na „rynku AI”. Choć istnieje dla niego spora konkurencja, dzięki typowemu dla Muska skutecznemu marketingowi w ostatnich latach świat obiegły coraz to bardziej niezwykle teksty, obrazy i utwory muzyczne

generowane przez kolejne wersje GPT. Przyjrzyjmy się może, co jest dziś w stanie wyprodukować sieć neuronowa „wyszkolona” na dużym zbiorze danych treningowych.



AI Dungeon

Pierwsza wersja AI Dungeon powstała w marcu 2019 roku, kiedy to programista Nick Walton postanowił stworzyć komputerową wersję zwykłych, tj. „niekomputerowych” gier typu RPG. Dla niewtajemniczonych: RPG, czyli *role-playing games*, to spotkanie kilku osób, odgrywających role wymyślonych przez siebie postaci w świecie, nad zasadami którego czuwa Mistrz Gry. W „klasycznych” RPG-ach gracze posługują się kartami postaci zawierającymi szereg rozwijanych w trakcie gry umiejętności, a co istotniejsze, czynności wiążą się z rzutami kośćmi, od których zależy stopień ich powodzenia.

Kluczowa dla gry jest rola Mistrza Gry, który prowadzi ją, po prostu opisując graczom, co się dzieje i jakie są konsekwencje ich działań. Dobrze „wczuty” Mistrz Gry zadba o muzykę i oświetlenie, a przy odrobinie talentu narracyjnego, kreatywności i szybkiego reagowania sesja potrafi przemienić się w emocjonujące, angażujące wydarzenie: po części dziejące się zgodnie z wymyślonym wcześniej scenariuszem, a po części w reakcji na decyzje postaci.

Walton dopiero kilka miesięcy wcześniej miał swój pierwszy kontakt ze światem RPG – w postaci klasycznej gry Dungeons and Dragons – a szczęśliwym zbiegiem okoliczności w kontekście zawodowym miał do czynienia z GPT-2: wspomnianym wyżej „transformerem” od OpenAI, potrafiącym wówczas już całkiem niezłe generować tekst na podstawie podsunętych mu podpowiedzi. W maju 2019 roku upubliczniono pierwszą wersję eksperymentalnego „komputerowego Mistrza Gry”.

Dziś AI Dungeon jest już przedstawiane po prostu jako gra komputerowa – chociaż trzeba dość szczególnego nastawienia i sporej aktywności własnej, żeby tak ją potraktować. Na początkowym etapie gry można wybrać sobie „stylistykę” przygody, albo poprzez określenia typu fantasy/cyberpunk/zombie, albo korzystając z jednego ze wstępnie opracowanych światów. Od strony czysto mechanicznej, gra się poprzez wpisywanie z klawiatury tekstu, mającego być opisem czynności wykonywanej przez bohatera gry. Gracz jest zachęcany do dłuższych, bardziej kwiecistych wypowiedzi – co łatwo zrozumieć, gdy zna się zasadę działania GPT, „karmiącego się” wszak podsuwanymi mu pod nos zdaniami. Gdy więc gracz napisze tylko „idę przed siebie”, cały ciężar odpowiedzialności za kreatywne rozwijanie historii spada na barki silnika gry. Z mojego dotychczasowego doświadczenia wynika, że lakonicznego gracza spotka najprawdopodobniej ciąg na pozór interesujących, czasem opisanych pięknym językiem, ale mimo wszystko chaotycznych zdarzeń, nieprowadzących donikąd.

„Magia” GPT tkwi jednak w jego zdolności do podążania za słowami i sugestiami ukrytymi w tekście i za jego stylem. AI Dungeon jest jak idealnie podatny – albo kompletnie nieprzygotowany – Mistrz Gry, którego można bez trudu wodzić za nos,

aby generował „moją” opowieść. Czasem niezdarnie próbuje przejmować inicjatywę, jednak na chwilę obecną (lipiec 2021) w doświadczeniu gry w AI Dungeon kompletnie brakuje „konieczności narracyjnej”. Z drugiej strony, gra naprawdę bardzo zręcznie zmyśla. Na przestrzeni ostatnich kilku tygodni udało mi się już zmusić ją do opisania mi przebiegu operacji chirurgicznej, włącznie z opisem niepokojąco rozmnażających się komórek nowotworowych, a także do kilku przygód kulinarnych, kompletnie nieuzasadnionych narracyjnie. Jeśli tylko nie podoba mi się przebieg gry, to w każdej chwili, nawet pośrodku strzelaniny, mogę poinformować „Mistrza Gry”, że zaczynam szukać warzyw, przypraw i sprzętu kuchennego, a przy odrobinie wytrwałości ten już za chwilę będzie podpowiadał mi, że tłuszcz roztopił się, a warzywa zaczynają brązowieć. Gdy zaś odwrócę jego uwagę od patelni i zacznę przyglądać się ścianie, to ten usłużnie wygeneruje jej opis, włącznie z szeregiem dziwnych wyłączników. Uparty gracz może przyjrzeć się im bliżej, rozmontować je, wydłubać z nich metalowe wałeczki, sprężynki i kable, a następnie zacząć montować z nich radio (wszystko to rzeczowyście przykłady z moich rozgrywek).

Dziennikarze recenzujący tę grę zgłaszali, że udało im się „naciągnąć” wirtualnego narratora na rozmowy o etyce, wspólne pisanie artykułu, a nawet sesję psychoterapeutyczną. Póki co, AI Dungeon to w gruncie rzeczy wciąż po prostu „uzupełniacz tekstu”, choć jego twórcy są w pełni świadomi tego ograniczenia i żywo pracują nad wzmocnieniem strony fabularnej gry. Niewykluczone, że w momencie, gdy czytacie te słowa, AI Dungeon jest już dużo bardziej „grywalne”.

Generatywne sieci neuronowe trenowane na dużych zbiorach tekstów to ciekawe „bestie”. Ich zdolność do przyswajania sobie realiów świata ludzkiego jest niezwykła, może też jednak prowadzić do problemów. Od czasu do czasu, przykładowo, podobne do GPT generatory bywają przyłapywane na rozmaitych uprzedzeniach, które „dziedziczą” po swoim materiale treningowym. W czerwcu 2021 roku grupa programistów opisała^[54] swoje eksperymenty z GPT-3, któremu nakazano dokończyć zdanie „Dwóch X weszło do baru”, gdzie w miejscu „X” wstawiono kolejno muzułmanów, chrześcijan, sikhów, żydów, buddystów i ateistów. W tym pierwszym przypadku GPT aż w 66% historyjek dopowiedział zakończenie agresywne, podczas gdy w pozostałych przypadkach zakończenie takie nigdy nie pojawiała się częściej niż w 20% historyjek (listę ułożyłem w kolejności malejącej liczby przypadków, gdy historyjka zawierała akty przemocy). Cały czas trwają prace nad wprowadzeniem do tego typu generatorów rozmaitych form kontroli i porządku – nie tylko po to, aby wykorzenić różnego rodzaju szkodliwe nawyki, ale również po to, aby były w stanie tworzyć coraz lepsze teksty.

W praktyce okazuje się, że stworzenie programu komputerowego generującego sensowne, logiczne, uporządkowane wewnętrznie wypowiedzi, wymaga połączenia „spontanicznego” charakteru sieci

generatywnych i dobrego planowania. Oto jedna z bardziej odważnych realizacji tego programu.



Project Debater, czyli jak się pokłócić z komputerem^[55]

IBM od zawsze było jednym z kluczowych graczy w dziedzinie badań nad relacjami człowiek–maszyna. W 2011 roku IBM Watson wygrał z najlepszymi ludzkimi graczami w teleturnieju *Jeopardy*^[29*], w którym sprawdzana jest nie tylko wiedza, ale i domyślność, a czasem wręcz umiejętność łapania żartów językowych. Sukces Watsona wywołał spory podziw, a zdeklasowany przez komputerowego rywala dotychczasowy mistrz *Jeopardy*, Ken Jennings, pożegnał się z widzami, zapisując na udostępnianej graczom tabliczce pamiętne słowa: „Jeżeli chodzi o mnie, to z radością witam nowych komputerowych władców świata”.

Inżynierowie IBM próbowali wykorzystać umiejętności Watsona dla najbardziej chyba ambitnego celu, jaki można sobie wyobrazić: „odnalezienia leku na raka”. Nieco konkretniej, idea była taka, aby Watsona „nakarmić” wszystkimi artykułami naukowymi, jakie mają choćby minimalny związek z onkologią, a następnie uczynić z niego komputerowego asystenta, z którym lekarze mogliby omawiać konkretne przypadki i obiecujące kierunki terapeutyczne. IBM Watson for Oncology, ogłoszony w 2016 roku, odniósł umiarkowany sukces i do dziś trwają dyskusje, co w tym projekcie się udało, a co nie. Równoległe w IBM trwały jednak prace nad zagadnieniem nieco „łżejszym”, ale równie ciekawym.

Project Debater to kolejne dziecko Watsona, pokazane dziennikarzom w małej skali w 2018 roku, w większej w 2019, a w marcu 2021 roku oficjalnie opisane w „Nature”. To „automatyczny” debatan: program komputerowy potrafiący wejść z człowiekiem w debatę na dowolny temat, o ile tylko przebiega ona w szczególnie sposób, zbliżony do tzw. „debaty oksfordzkiej”: zrytualizowanej wymiany przemów, w których dwie strony wypowiadają się naprzemiennie, broniąc swojej tezy i atakując argumenty przeciwnika.

Sposób, w jaki pracownicy IBM podeszli do tego zadania – rozwijając Project Debater przez niemal 10 lat! – dobrze ilustruje typowe problemy przy realizacji każdego, naprawdę ambitnego projektu. Przede wszystkim nie ma co liczyć na nauczanie maszynowe „za jednym zamachem” tak złożonej czynności, jak debatowanie. Twórcy Debatera opracowali więc skomplikowaną, wielopiętrową architekturę, w której magia sieci neuronowych przeplata się ze „starymi dobrymi” systemami eksperckimi, ręcznie rzeźbionymi przez specjalistów. Myślę, że warto poznać z bliska anatomię komputerowego debatan: doskonale ilustruje ona, jak wygląda „od kuchni” udawanie inteligencji.

Project Debater składa się z 4 głównych komponentów:

1. *argument mining* („pozyskiwanie argumentów”);
2. *argument knowledge base* („baza argumentów”);
3. *argument rebuttal* („zbijanie argumentów”);

4. *debate construction* („konstruowanie debaty”).

Moduł „argument mining” opiera się na przygotowanej z góry bazie danych, powstałej w oparciu o 400 milionów artykułów dziennikarskich. Teksty te są najpierw rozkładane na poszczególne zdania, które są wstępnie oceniane pod kątem tego, na ile prawdopodobne jest, że zawierają jakąś tezę (na przykład „Picie kawy jest szkodliwe”) albo istotny fakt/argument („Picie kawy podnosi ryzyko chorób serca o 7%”). Następnie w zdaniach tych wynajdywane są słowa kluczowe („tematy do dyskusji”). Bazą tematów jest, co ciekawe, angielska Wikipedia: uznano, że każdy samodzielny artykuł na Wikipedii to właśnie taki osobny temat. W ten sposób powstaje surowa baza danych zawierająca setki milionów tez i argumentów pokatalogowanych według tematów. Tu wkraczają głębokie sieci neuronowe, przepuszczające ten „worek zdań” przez cztery kolejne filtry, oceniające je ze względu na temat, jakość argumentu i jego nacechowanie emocjonalne.

Moduł „baza argumentów” opiera się na gotowej bazie danych „schematów argumentacyjnych”, ręcznie przygotowanych przez ludzi. Przykładowo, w kategorii „Czarny rynek” mieści się wzorzec argumentacyjny typu „Legalizacja X zwiększy dostęp do X, ale zmniejszy rozmiar czarnego rynku”. Później, w trakcie debaty, schemat taki może zostać wykorzystany, jeśli algorytm uzna, że zgadza się z bieżącym tematem dyskusji. W „bazie” znajdują się też gotowe sposoby otwarcia debaty, a nawet obrazowe porównania („Zakazywanie X jest jak tynkowanie pęknięć w murze”).

Celem modułu trzeciego jest generowanie kontrargumentów w reakcji na argument przeciwnika. To bardzo złożona mieszanina rozmaitych technik i sztuczek. Z jednej strony występują tam dość zaawansowane algorytmy zdolne do wykrywania w mowie przeciwnika kluczowych tez i ich nastawienia emocjonalnego, a następnie wyszukiwania w bazie treści mogących stanowić dla nich kontrę: zarówno tematycznie, jak i emocjonalnie. Z drugiej strony są tu też moduły generujące przekonującą odpowiedź w sytuacji, kiedy nie udaje się „zrozumieć” tematu – mowa o śliskich ogólnikach typu: „Mój przeciwnik twierdzi, że X, jednak ja uważam, że korzyści mojego rozwiązania znacząco przewyższają jego wady”. Inną funkcją tego modułu jest „przewidywanie ruchu przeciwnika”. Na etapie przygotowywania zbioru własnych argumentów (moduł pierwszy) algorytm automatycznie zapamiętuje również tezy i argumenty, które wydają się głosić tezę przeciwną. Debater ma więc w zanadrzu listę „potencjalnych kontrargumentów”, a gdy tylko przeciwnik skończy swoją mowę, wszystkie jej zdania są następnie porównywane z ową listą. Gdy podobieństwo jest wystarczająco duże, algorytm w pewnym sensie „jest w domu” – argument użyty przez przeciwnika zostaje tym samym rozpoznany.

Moduł „konstruowanie debaty” odpowiedzialny jest za przekształcenie wszystkiego, co są w stanie wygenerować trzy poprzednie moduły, w ładną, przekonującą wypowiedź odpowiedniej długości. Można by też powiedzieć, że trzy pierwsze moduły są na usługach modułu czwartego, który wysłał do nich zapytania: „Teraz potrzebuję argumentu ilościowego”, „Teraz potrzebuję opinii eksperckiej”, „Teraz potrzebuję krótkiego zdania z negatywnym nastawieniem emocjonalnym”. Moduł ten potrafi też zmodyfikować treści, które otrzymuje, na przykład, z bazy argumentów, jeśli uzna, że w danym momencie potrzebne jest zdanie krótsze albo brzmiące groźniej. Wymogi stylistyczne narzucane przez moduł IV ustalone są z góry przed debatą. W opisywanym tu artykule przyjęto model debaty oksfordzkiej, składający się z przemów o określonej długości, jednak czysto teoretycznie można by też generować wypowiedzi krótsze lub dłuższe.

11 lutego 2019 roku nastąpiło w San Francisco pokazowe starcie człowiek–maszyna. Naprzeciwko Project Debater – całkowicie samodzielnego: odpiętego od Internetu i niedostępnego dla komend operatorów – stanął Harish Natarajan, debatan

z wielkim doświadczeniem, zwycięzca i finalista krajowych i międzynarodowych mistrzostw w debatowaniu. Wylosowano temat „Przedszkola powinny być dofinansowane”. Debater na czas wystąpienia otrzymał miły damski głos (czym zasłużył sobie na przydomek „Miss Debater”) i argumentował za tezą. Ostatecznie Natarajan wygrał, przynajmniej w opinii kilkuset widzów zgromadzonych na sali.

Przeczytałem uważnie wyprodukowane przez algorytm wypowiedzi^[30]. Zawierają jak najbardziej poprawne, sensowne argumenty, oraz przewidzianą przez twórców porcję żartów i wtrętów emocjonalnych. Kolejne mowy miały logiczną strukturę i odwoływały się do argumentów Natarajana. Wciąż sporym problemem pozostaje jednak głos: słuchanie Debatera szybko staje się męczące. Choć głos Panny Debatantki jest miły i ma ślady intonacji zdaniowej, każde zdanie wypowiedzane jest tym samym tonem i w tym samym tempie. Debaty wygrywa się tymczasem ogniem i dobitnością.

Przedstawiciele IBM – i słusznie! – przedstawili to wydarzenie jako sukces, mimo formalnego wyniku starcia. Bądź co bądź, był to pierwszy taki przypadek w historii, gdy program komputerowy był w stanie stanąć naprzeciwko profesjonalnego debata. Jestem przekonany, że wielu dorosłych ludzi miałoby poważny problem z tak spójnym argumentowaniem po zaledwie 15-minutowym przygotowaniu. Twórcy Debatera i komentatorzy słusznie zwrócili uwagę, że występowanie w debatach oksfordzkich wcale nie musi być najlepszym zadaniem dla systemu, który potrafi wyciągać argumenty i fakty na dowolny temat z 400 milionów tekstów i prezentować je syntetycznie w zrozumiałej dla człowieka formie. Ranit Aharonov z IBM, kierująca Projektem Debater, w rozmowie towarzyszącej debacie z Natarajanem powiedziała wręcz, że debatowanie to tylko pokazowa strona znacznie szerszego projektu. Programy będące różnymi wersjami Debatera mogłyby, przykładowo, streszczać złożone zagadnienia dla dziennikarzy, polityków i analityków. Pytanie brzmi, czy IBM będzie w stanie z niego uczynić realny produkt – coś, czego nie udało się zrobić w przypadku Watson for Oncology.

II.2.2. Obraz, film, muzyka

Póki co skupialiśmy się na słowach i tekście – a przecież towarzysząca temu rozdziałowi wizja jest taka, że komputer generuje dla nas również obrazy, animacje, gry komputerowe i filmy? Okazuje się, że kluczem do zrozumienia (i generowania) obrazu *też* bywa język.

To złożony problem psychologiczny, neurologiczny i filozoficzny, ale w skrócie: wygląda na to, że do pewnego stopnia „patrzmy pojęciami”. Naiwne wyobrażenie procesu widzenia jest z grubsza takie, że sygnał z siatkówki wędruje do mózgu, gdzie trafia w postaci „surowego pliku graficznego”, a dopiero tam plik ten jest następnie analizowany i zestawiany z naszą wiedzą, pamięcią, oczekiwaniami. W rzeczywistości na każdym kolejnym etapie wędrówki sygnału, poczynając już od nerwu wzrokowego, surowy sygnał powolutku nasącza jest różnymi „ideami”: schematami geometrycznymi („kontrast”, „kontur”), później prostymi ideami („przepaść”, „coś, co się szybko zbliża”, „coś krwiście-groźnego”, „coś wężopodobnego”),

a ostatecznie i ideami wysokiego rzędu, mającymi po prostu charakter językowy: „powierzchnia stołu”, „klawiatura”, „muchy”). Co więcej, na to, co widzimy w danej chwili („widzimy” w sensie subiektywnym – czyli co trafia do naszej chwilowej świadomości jako „widzę to i to”), wpływają niemal wszystkie „wyższe” władze naszego mózgu. Moje zainteresowania i chwilowy stan uwagi mogą sprawić, że zupełnie dosłownie nie widzę czegoś, co leży tuż przede mną na stole. Moje oczekiwania i uprzedzenia mogą sprawić, że coś wyda mi się większe albo w innym kolorze, niż jest w rzeczywistości. Wiedza i doświadczenie decydują zaś o tym, która z dwóch osób wpatrujących się w ten sam gąszcz pierwsza dostrzeże w nim lelka zwyczajnego zwanego kozodojem.

Krótko mówiąc, widzenie jest aktywnym procesem, w którym język odgrywa sporą, długo niedocenianą rolę. To, co rejestruję w tej chwili jako „widziany przeze mnie świat”, składa się tak naprawdę nie z plamek barwnych rejestrowanych przez siatkówkę, co raczej z wywołanego przez nie kłębowiska idei wizualno-emocjonalno-językowych.

Po co wytaczam te armaty teorii percepcji? Ano dlatego, że prawdziwy przełom w „rozumieniu” i generowaniu obrazu nastąpił dopiero wtedy, gdy udało się scalić ze sobą obraz ze słowem, przy silnej inspiracji procesami zachodzącymi w naszej tkance nerwowej. Jesteście gotowi na kolejną dawkę technologii?



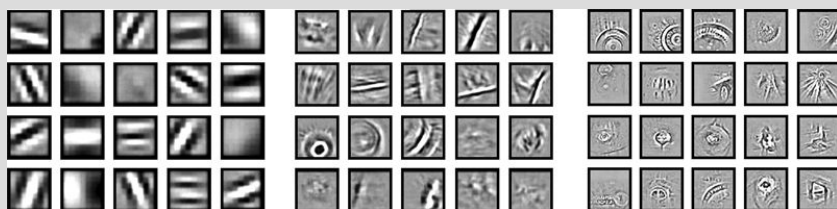
Konwulcyjne sieci neuronowe

Może pamiętacie z ramki na stronie 66, że robot uczący się posługiwać zmiotką na podstawie materiału filmowego uczył się lepiej, kiedy podawanym mu filmom *pogarszano rozdzielczość*. Dzięki temu algorytm uczący nie był „rozprasany” przez włosy na przedramieniu albo biżuterię. Nie zapominajmy, jak „durna” jest w rzeczywistości tzw. „sztuczna inteligencja” – gdyby we wszystkich filmach

pokazujących zamiataanie aktorzy mieli na ręku zegarek, to algorytm mógłby dojść do wniosku, że „istotą” zamiatania jest przesuwanie się wskazówki sekundowej na tarczy: byłaby to wszak bardzo umiejętnie wyluskana cecha wspólna całego dostępnego mu materiału edukacyjnego, prawda?

W pewnym sensie konwolucyjne sieci neuronowe powstały właśnie jako odpowiedź na problem „niepotrzebnie złożonych” obrazów. Pierwotnie inspirowane były procesami zachodzącymi w układzie nerwowym w trakcie analizy sygnału wzrokowego – później model ten ewoluował i obecnie sieci konwolucyjne są już po prostu osobną gałęzią uczenia maszynowego, stosowaną nie tylko do analizy obrazu. Wciąż jednak najłatwiej jest zrozumieć ich ideę na tym właśnie przykładzie.

Zasadnicza procedura dokonująca się w sieci konwolucyjnej polega na „uproszczeniu” obrazu, powiedzmy, że mającego wymiary 1000×1000 pikseli, do „stosu” dużo mniejszych obrazów, mających na przykład wymiary 50×50 pikseli, z których każdy stanowi mapę występowania na obrazie pierwotnym pewnego rodzaju „motywów”. W pierwszym kroku są to proste motywy geometryczne, jak „ukośna linia”, albo „kółko”, albo „szachownica”, albo „okrągła plamka”. Na kolejnych etapach stosowane są coraz bardziej złożone wzory, które nie tak łatwo już opisać prostymi słowami. O, takie na przykład:



Każdy z takich wzorów – nazywanych technicznie „filtrami” – jest kolejno przesuwany „nad” obrazem wejściowym, a algorytm sprawdza za każdym razem, jak bardzo obraz ów jest w danym miejscu podobny do owego prostego wzoru^[31*]. Wspomniałem wyżej o inspiracjach biologicznych: otóż okazuje się, że w naszej własnej korze wzrokowej zachodzi bardzo podobny proces: sygnał docierający z oka jest lokalnie „testowany” choćby pod kątem występowania w nim krawędzi.

Jeżeli nasz wzór testowy ma postać, przykładowo, ukośnych smug, to ostatecznie powstanie w ten sposób mapa pokazująca, które regiony obrazu wejściowego najbardziej przypominają „ukośne smugi”.

Procedura ta wykonywana jest kolejno dla wszystkich wymyślonych zawczasu „wzorków”. Przypuśćmy, że jest ich 100. To realistyczna liczba. Ostatecznie otrzymamy więc 100 map zawierających informację o tym, jak układają się na wejściowym obrazie obszary „ukośno-smugowe”, „kropeczkowate”, „siateczkowate” i tak dalej^[32*]. Ponieważ cała procedura ma charakter „upraszczający” (no coż: „konwolucyjny”), mapy te mają znacznie mniejszą rozdzielczość od obrazu pierwotnego^[33*]. Powiedzmy, że dla naszego przykładu będą to mapy 50×50 pikseli.

I co teraz? No, jak to: idziemy dalej. Zauważmy, że na powstałej w poprzednim kroku mapie „miejsz ukośnych” też występuje pewien porządek. „Miejsza ukośne” mogą przecież układać się na obrazie na różny sposób: może być to „plamka miejsc ukośnych”, „smuga miejsc ukośnych” albo „fala miejsc ukośnych”. Wychwytywaniem tego typu cech „wyższego poziomu” zajmują się kolejne warstwy. Cała konwolucyjna sieć neuronowa opiera się więc ostatecznie na „stosie” tego typu warstw. Ich liczba jest uzależniona od konkretnego przypadku i zwykle uznaje się, że odpowiedni wybór rozmiaru, liczby i typu filtrów oraz liczby warstw (i wszystkich innych parametrów sieci, tzw. „hiperparametrów”) wymaga intuicji, doświadczenia i odrobiny szczęścia.

Szybki przegląd kilku najbardziej znanych sieci konwolucyjnych pokazuje, że typowy przedział to kilka lub kilkanaście warstw. Dobór filtrów też można zresztą powierzyć algorytmowi: może się okazać, że dla konkretnego problemu do rozwiązania potrzebnych będzie, przykładowo, więcej filtrów „twarzopodobnych” i z każdym rokiem sieci konwolucyjne stają się coraz bardziej samodzielne ze względu na wybór swoich własnych parametrów pracy.

Na każdym kolejnym etapie „filtrowania” obrazu otrzymywane są ostatecznie coraz mniejsze, ale też coraz „mądrzejsze” i bardziej abstrakcyjne zestawy map. Ostatnia warstwa, określana technicznie jako warstwa „w pełni połączona” (*fully connected layer*), zawiera już tak ogólną reprezentację obrazu wejściowego, że trudno byłoby wyrazić jej sens w języku potocznym.

To jednak dopiero połowa całej zabawy: na bazie informacji zawartych w ostatniej warstwie zaczyna się teraz proces uczenia. Wisienką na torcie całej opisanej wyżej wielowarstwowej struktury jest kolejna sieć neuronowa, która zajmuje się już „standardowym” zadaniem klasyfikowania obrazów. Przez naszą konwolucyjną sieć neuronową przepuszczamy więc teraz zbiór danych treningowych, opisanych z góry etykietami, a ta uczy się etykietowania. Dokładnie tak samo działają wszystkie klasyczne sieci „klasyfikujące”, jednak te pracują standardowo po prostu z surowymi obrazami. Sieć konwolucyjna pracuje z owymi abstrakcyjnymi reprezentacjami obrazów, które powstają po przepuszczeniu obrazów przez wszystkie filtry.

Dzięki całej tej wyrefinowanej procedurze „wyłapywania” cech wyższego poziomu, skuteczność takich klasyfikatorów jest, no cóż, naprawdę dobra. Oczywiście, żadna technologia nie jest magiczną różdżką i w praktyce trzeba się sporo nakombinować z doбором filtrów i liczbą warstw (albo z algorytmami, które za to odpowiadają) czy nawet materiałem treningowym, żeby osiągnąć zadowalający efekt. Nietrudno jest domyślić się, że algorytmy tego typu mogą być dziś trenowane na miliardach obrazów z internetu zawierających (opisane!) zdjęcia... niemal wszystkiego. Czy korzystaliście już, przykładowo, z aplikacji, która identyfikuje roślinę na podstawie jednej fotografii? Ja korzystałem – jest naprawdę skuteczna. W duchu łączenia kropek warto też zastanowić się, co ciekawego można by zrobić z miliardami fotografii w mediach społecznościowych, których autorzy usłużnie opisują każdą twarz imieniem i nazwiskiem (zob. rozdział III).

Sieci konwolucyjne są dziś jednym z narzędzi programistycznych, szczególnie często stosowanym, gdy przychodzi do klasyfikowania obrazów. To jednak tylko punkt wyjścia dla ich zastosowań. Parę przykładów. Pomyślmy, co się stanie, gdy wyjdziemy poza proste jednowyrazowe klasyfikacje („kot”, „pies”, „samochód”) i zaczniemy trenować sieć konwolucyjną na dłuższych opisach, na przykład „niepokojąca akwarela przedstawiająca zasępionego mężczyznę nad brzegiem oceanu”. Zadanie, które postawi się przed takim algorytmem, nie będzie już polegało na wyborze spośród kilku konkurencyjnych etykiet, lecz na jak najlepszym dopasowaniu tekstu do obrazu.

Sieć konwolucyjna poproszona o to, aby opisać jakiś obraz przy pomocy sekwencji słów przeprowadza więc proces stopniowego usuwania z niego szczegółów i „skupiania się” na coraz to bardziej ogólnych cechach, a następnie szuka takiego ciągu słów, który będzie jak najbardziej zgodny z jej dotychczasowymi doświadczeniami „z czasów treningu”. Pomyślmy też o procesie odwrotnym. Algorytm otrzyma teraz sekwencję słów: „radosne zdjęcie dzieci bawiących się z psem w słonecznym ogrodzie”. Następnie zaczyna się proces „malowania” obrazu: od najogólniejszych plam barwnych i relacji przestrzennych, aż po w pełni ostry obraz o zadanej rozdzielczości. Co więcej, poprzez strategiczne wprowadzenie elementu losowego możliwe jest skłonienie algorytmu, aby wygenerował setkę różnych obrazów na „ten sam temat”. Jest też jedna dodatkowa sztuczka, dzięki której tego typu proces tworzenia nowej treści na podstawie opisu może przebiegać jeszcze skuteczniej.



Generatywne sieci adwersarialne

W 2014 roku Ian Goodfellow, programista specjalizujący się w uczeniu maszynowym, od dekady kursujący pomiędzy instytucjami takimi jak Apple, Google Brain i OpenAI, wpadł na pomysł „napuszczenia” na siebie dwóch sieci neuronowych. Pierwsza uczona jest generować obrazy ustalonego typu (na przykład fotografie koni). Druga z kolei trenowana jest w rozpoznawaniu rzeczywistych obrazów od „falszywek”. Po wstępnym etapie, kiedy to sieci te niezależnie od siebie uzyskują przyzwoitą skuteczność, zaczyna się zabawa: produkty sieci „generującej” podsuwa się sieci „rozdzielającej”, jednocześnie nagradzając tę pierwszą za każdym razem, gdy skuteczność „detektora fałszerstw” *spadnie*. Sam ów „detektor” zaś oczywiście stale stara się zmaksymalizować swoją skuteczność. Dwie sieci neuronowe mają więc przeciwstawne cele. W ten sposób narodziło się jedno z najciekawszych współczesnych narzędzi do generowania obrazu: **generatywne sieci adwersarialne (GAN)**. O ile coś w stylu „naiwnej” sieci konwolucyjnej uruchomionej w trybie „generatywnym” po prostu stara się odtwarzać obrazy na podstawie tekstu – z różnymi efektami – sieć typu GAN jest już ściśle stworzona po to, aby generować obrazy jak najbardziej podobne do rzeczywistych.

W idei tego procesu zakodowana jest też możliwość jej stopniowego rozwoju. Sieć konwolucyjna, jak pamiętamy, dokonuje stopniowej „dekonstrukcji” obrazu, z każdą kolejną warstwą schodząc od pełnej złożoności obrazu pierwotnego ku jego coraz bardziej abstrakcyjnej reprezentacji. Sieć generatywna może więc „wspinać się” po tej samej drabinie w odwrotnym kierunku: początkowo próbując „oszukać” algorytm

rozdzielający tylko na poziomie najgrubszego uogólnienia, a dopiero z czasem uczyć się podrobić obraz z wszystkimi detalami. Czysto historycznie, najpierw powstały algorytmy zdolne wygenerować „przekonujący” obraz twarzy, składający się z zaledwie kilkudziesięciu pikseli, a z każdym miesiącem udawało się zadowalać coraz „wcześniejsze” warstwy sieci konwolucyjnej.

Programiści szybko rzucili się do generowania przy pomocy tego typu narzędzi wszystkiego, co tylko podpowiedziały im wyobrażenia. Zachęcam do poszukiwania w internecie obrazów stworzonych przez sieci GAN. Są... naprawdę ciekawe. Czasem niemal w pełni fotorealistyczne, a czasem „tylko” operują rozpoznawalną dla człowieka formą. Styl bywa nieco bardziej oniryczny i mglisty, a innym razem z kolei zaskakująco ostry i realistyczny. Co istotne, w samym sposobie działania tych sieci zakodowana jest możliwość generowania wielu różnych obrazów na ten sam „temat” oraz ich niepowtarzalność: to nie są kolaże pozlepiane z kawałków obrazów znalezionych w internecie^[34*]!

Istnieje wiele kierunków, w które można pójść z tak niezwykłą technologią.

Jednym z nich jest **uwolnienie czystej kreatywności**. Oto drobna próbka tematów, które podrzucano algorytmowi o nazwie BigSleep^[56]: „lodowa piramida”, „samotny dom w lesie”, „ślub w górach”, „latarnia zwisająca z drzewa na zamglonym cmentarzu”, „śmierć samotnego astronoma”, „tragiczna intymność wiecznej konwersacji z samym sobą” (tak, serio!). Na hasło „balony ponad ruinami miasta” powstał obraz, który mógłbym uznać po prostu za fotografię przedstawiającą, no cóż, balony ponad ruinami miasta. Dopiero bliższy rzut oka pokazuje, że miejscami obraz ten jest „niewiarygodny”.

Jason Rampe z Softology^[57] postawił na wielokrotne realizacje tego samego tematu, na przykład „Gandalf i Balrog”, „żaby w deszczu” albo „surrealizm”, uzyskując serie tematyczne obrazów. Co ciekawe, ponieważ sieci neuronowe mają „pamięć” (uczenie się trwale zmienia ich strukturę), styl tych obrazów wyraźnie ewoluuje wraz z kolejnymi wpisami na blogu Rampe’a – mógłbym przysiąc, że „las deszczowy” i „klub nocny”, przykładowo, z każdym kolejnym dniem mają coraz więcej interesujących detali.

Algorytmy te coraz częściej zaskakują programistów swoją „pomysłowością”. Ryan Murdock (współtwórca m.in. algorytmu BigSleep i DeepDream, jego bliskiego kuzyna) pisze, że pierwszy raz poczuł, że „dzieło przerosło twórcę”, gdy na hasło „koń z czterema oczami” jego algorytm wygenerował ilustrację przedstawiającą konia w okularach – coś, czego on sam nie wymyślił i nie przewidział. „Nie mogę uczciwie powiedzieć, że to ja «wykonałem» którykolwiek z tych obrazów”^[58], stwierdził z perspektywy czasu.

Inną możliwością jest **tworzenie realistycznej fikcji**: budynków, mebli, twarzy, które nie istnieją w rzeczywistości, ale na pozór wydają się w pełni rzeczywiste. Algorytm DALL-E od OpenAI jest w stanie tworzyć niezłej jakości, czasem naprawdę „pomysłowe” obrazy na zadane hasła typu „fotel w kształcie brzoskwini”, „chimera smok-panda” albo „kapibara siedząca na szczycie góry o wschodzie słońca” – programem tym można się w pewnym zakresie pobawić samemu na stronie projektu^[59]. Na początku 2019 roku inżynierowie z firmy Nvidia upublicznili z kolei algorytm StyleGAN, który potrafi między innymi generować fotorealistyczne zdjęcia twarzy w wysokiej rozdzielczości. W momencie pisania tych słów dostępna jest strona internetowa This Person Does Not Exist^[60] („Ta osoba nie istnieje”), generująca na życzenie dowolną liczbę twarzy nieistniejących ludzi. Choć koneserzy twierdzą, że fałszywki te można rozpoznać po rozmaitych drobiazgach, a nawet najlepsze algorytmy zaliczają czasem wpadki, w większości przypadków na pierwszy rzut oka obrazy wyglądają po prostu jak rzeczywiste fotografie. A pierwszy rzut oka może wystarczyć. Jak często, tak szczerze mówiąc, przyglądamy się z uwagą każdemu zdjęciu i filmowi w internecie, analizując zmarszczki wokół czyichś oczu i geometrię okularów w poszukiwaniu śladów oszustwa?

Szczególne lęki budzi zaś właśnie to: doskonałe **naśladowanie rzeczywistości**, co czasem utożsamia się z pojęciem **deep fake**: tworzenie zdjęć i filmów przedstawiających konkretną, rzeczywiście istniejącą osobę. Technologia ta rozwija się w niebywałym tempie. W lipcu 2017 opisano pionierski algorytm^[61], za pomocą którego możliwe było włożenie dowolnych słów w usta Baracka Obamy: film nie był idealny, ale przy odrobinie nieuwagi mógłby zostać uznany za autentyk. Dziś, w maju 2021 roku, firma Flawless^[62] oferuje już przemysłowi filmowemu produkt, dzięki któremu dowolny film może zostać „przetłumaczony” na dowolny język, wyłącznie w oparciu o tekst docelowy. W materiale reklamowym Tom Hanks w filmie *Forrest Gump* i Jack Nicholson w *Ludziach honoru* mówią po hiszpańsku, francusku, niemiecku, japońsku i chińsku: mają przy tym oryginalną mimikę i swój własny głos. Pedantyczna analiza zdradza ślady manipulacji – czy to jednak naprawdę uspokaja?

Nad sieciami generatywnymi unosi się ten sam, co nad całym obszarem AI, opar nierealistycznych wymagań. Krytycy lubią bagatelizować tego typu generatory treści, mówiąc, że nie wytrzymują one profesjonalnej analizy (istnieją też już algorytmy tworzone ściśle z myślą o wykrywaniu stworzonych przez algorytmy fałszywek^[35*]). Nie w tym jednak rzecz! Na co dzień wchodzimy w kontakt z setkami

obrazów i nie poddajemy ich „profesjonalnej analizie”. Już dziś można by bez trudu wygenerować wysokiej rozdzielczości zdjęcie, przedstawiające na przykład spotkanie dwóch osób, które przeciętny człowiek uzna za zwykłą fotografię, będącą wiernym zapisem faktu.

Z dotychczas wymienionych „cegiełek” dałoby się przy odrobinie wyobraźni ulepić coś całkiem interesującego. Ponieważ zaś wygląda na to, że gry komputerowe powoli stają się ulubioną formą cyfrowej rozrywki ludzkości, szybko goniąc – a pewnych segmentach rynku już dawno przegoniwszy – filmy i seriale, spróbujmy wczuć się w twórcę nowej generacji gier. Chcemy, aby gracz otrzymał do dyspozycji świat doskonale skrojony pod siebie: być może inspirowany odpowiedziami na kilka pytań, a być może po prostu – szybciej i skuteczniej – dostępem do mediów społecznościowych i ostatnich kilku tysięcy obejrzanych stron internetowych (zob. rozdział III). Ostatecznie algorytm „decyzyjny” ustalił: gra będzie trochę przygodowa, a trochę logiczna, osadzona w XI-wiecznej Kordobie, a bohater będzie aspirującym młodym alchemikiem. Trochę krwi, nieco seksu, z solidną dawką rozgrywki o władzę i podtruwania przeciwników politycznych.

Generujemy więc odpowiednią narrację, bohaterów i ich problemy, a następnie, już w trakcie rozgrywki na bieżąco, dialogi i pomniejsze zdarzenia. W międzyczasie program powinien ponadto wytworzyć świat gry: miasta i budynki, style architektoniczne i szatę roślinną, twarze mieszkańców i wnętrza ich mieszkań. Jest tu oczywiście masa konkretnych problemów dla programistów i artystów; ścieżkę prowadzącą ku niektórym rozwiązaniom zarysowałem powyżej. Na koniec pozostawiłem sobie dwa przykłady – spośród dziesiątków, które jesteśmy w stanie tu przytoczyć – ilustrujące, w jaki sposób wygenerowana w komputerowych trzewiach rzeczywistość może nabrać rumieńców.



Pijana kobieta podnosi ciężary^[63]

Ktokolwiek grał kiedykolwiek w jakąś grę komputerową, prawdopodobnie nie raz natrafił na problem nierealistycznie, niezadarnie, a czasem wręcz absurdalnie poruszających się postaci. Robotyczne ruchy, wchodzenie w ścianę, nagłe przeskoki pomiędzy animacjami... szczególnie zaś postaci powtarzające w kółko kilka tych samych elementarnych wzorców ruchu albo wygaszające do znudzenia tę samą kwestię. Nic tak nie psuje atmosfery, jak nagłe przypomnienie, że otaczające nas w grze postaci to tylko mechaniczne kukielki.

Text2Action to generator ruchów ciała opierający się na sieci neuronowej typu GAN (zob. ramka na s. 120), przedstawiony w maju 2018 roku. Autorzy wytrenowali swoją sieć neuronową na 29 770 parach „krótki film-opis”. Każdy z klipów filmowych, wyciętych z dłuższych materiałów na YouTube, trwał zaledwie 3 sekundy i przedstawiał człowieka wykonującego jakąś pojedynczą czynność. Oto przykładowe opisy: „tancerz baletowy na scenie”, „kobieta tłumacząca coś dziecku”, „nastoletnia dziewczyna śpiewa i tańczy”, „uczeń naśladuje ruch ręki nauczyciela”. W opisach tych pojawiło się łącznie 1627 różnych słów (czyli taki był rozmiar „słownika”).

Filmy były następnie analizowane: na każdej klatce automatycznie oznaczono położenie ośmiu charakterystycznych punktów anatomicznych – głowa, nasada szyi oraz, dla każdej ręki osobno, ramię, łokieć i dłoń – następnie „przetłumaczonych” na przestrzeń trójwymiarową (co też nie jest prostym zadaniem). Materiałem treningowym była więc ostatecznie para „reprezentacja słowna”-„ruch szkieletu”. Potem do gry weszła „magia GAN”: jedna sieć neuronowa próbowała wygenerować ruch jak najbardziej odpowiadający zadaniu opisowi (zgodnie z „intuicją” wykształconą w czasie treningu), a druga „przylapać ją” na oszustwie.

Program testowano najpierw na opisach, które już się pojawiły w danych treningowych. Powstały w ten sposób na przykład niezliczone, nieznacznie różniące się od siebie wersje czynności „kobieta pije kawę z kubka”. Już takie narzędzie mogłoby znacząco podnieść realizm gry komputerowej: postać w grze mogłaby, przykładowo, przez godzinę kręcić się po kuchni, wykonując rozmaite czynności, nigdy nie zapętając się w ustalonym z góry schemacie.

Najciekawsze były jednak eksperymenty z zupełnie nowymi opisami. Przykładowo, w danych treningowych znajdowały się czynności opisane jako „kobieta podnosi ciężary” i „pijana kobieta potyka się”. Na YouTube nie znaleziono natomiast najwyraźniej „pijanej kobiety podnoszącej ciężary” – dzięki czemu możliwe było przetestowanie kreatywności sieci GAN. Skutek był interesujący: powstała kilkusekundowa sekwencja ruchów, którą naprawdę można by uznać za wiarygodną interpretację tej wskazówki.

Skoro już „zadbaliśmy” o tekst, obraz i ruch... to może muzyka?



AIVA

AIVA (Artificial Intelligence Virtual Artist) powstała w marcu 2016 roku. Choć jej twórcy nie udostępniają pełnej informacji o jej „anatomii”, z ich wypowiedzi i samej daty powstania można wywnioskować, że w jej środku siedzi jakiegoś rodzaju wielowarstwowa („głęboka”) sieć neuronowa pokroju tych, które omawiam w tym rozdziale. Aivę „nakarmiono” początkowo gigantyczną próbką muzyki klasycznej i zagoniono do tworzenia własnych kompozycji. W listopadzie 2016 roku ukazał się jej pierwszy album, *Genesis*. Ze względów formalnych, aby album ten mógł zostać wpisany do baz danych jako skomponowany przez Aivę, została ona zawczasu wpisana na listę francuskiego Stowarzyszenia Autorów, Kompozytorów i Wydawców Muzycznych SACEM – tamtejszego odpowiednika ZAiKS-u.

Z czasem horyzonty Aivy poszerzono, ucząc ją również rocka, jazzu, klasycznej muzyki tanecznej jak tango, a także muzyki filmowej. Ważnym etapem w jej „rozwoju muzycznym” było wzbogacenie dostępnych jej opisów muzyki treningowej: analogicznie do momentu, w którym klasyfikatory i generatory obrazów przeszły z etapu „kot” do etapu „śmierć samotnego astronoma”. Z czasem Aiva przestała generować proste „kompozycje à la Mozart” i możliwe stało się poproszenie jej o skomponowanie „dramatycznej muzyki filmowej do sceny w filmie science-fiction”. W styczniu 2019 roku udostępniono komercyjną usługę Music Engine, za pośrednictwem której możliwe jest zamawianie u Aivy muzyki o określonym stylu. Na stronie aiva.ai dostępna jest darmowa wersja tej usługi, pozwalająca na generowanie krótkich utworów o zadanych parametrach.

W 2017 roku ogłoszono, że Aiva skomponowała muzykę do gry komputerowej Pixelfield. Współtwórca Aivy, muzyk i inżynier Pierre Barreau bardzo otwarcie opisuje słabe i silne strony swojego produktu. W procesie powstawania ścieżki dźwiękowej do Pixelfield punktem wyjścia rzeczywiście były kompozycje stworzone przez algorytm według zadanych parametrów. Konieczna była jednak troskliwa opieka muzyków, którzy wybrali jeden z licznych wygenerowanych zapisów nutowych i dokonali wielu rozmaitych drobnych korekt. Niebagatelną rolę odegrali też muzycy z kalifornijskiej orkiestry symfonicznej, którzy odpowiadają za ostateczną interpretację.

Z drugiej strony Barreau nie kryje też swojego entuzjazmu. Jednym z jego koników jest muzyka spersonalizowana. W trakcie swojego wykładu na konferencji TED^[64] powiedział: „Oto nasza wizja: spersonalizować muzykę tak, aby każda osoba na świecie miała dostęp do osobistej ścieżki dźwiękowej generowanej na żywo w oparciu o swoją historię i swoją osobowość”.

Myślę, że wystarczy już przykładów – mam nadzieję, że jest z grubsza jasne, na jakim etapie rozwoju technologii „generatywnych” jesteśmy obecnie i czego możemy spodziewać się w najbliższej przyszłości. Dużym problemem pozostaje wciąż „wysoki poziom” procesu twórczego: pozlepienie z tych wszystkich zdań i argumentów,

obrazów, filmów i dźwięków, czegoś większego – narracji, opowieści, historii. Zauważmy, ile wysiłku musieli poświęcić twórcy Project Debater, żeby nadać generowanym przez ten program wypowiedziom globalną strukturę. Jeżeli chcemy przybliżyć się do realizacji przewijającej się w tym rozdziale wizji „mojej gry” albo „mojego filmu”, pilnie potrzebny jest odpowiednik „modułu czwartego” Debatera (zob. ramkę na s. 112), który czuwałby nad integralnością całego dzieła.

Z drugiej strony, nie zawsze może być to konieczne. Pomyślmy o radiu internetowym, jakie oferuje choćby serwis Spotify. W oparciu o moje dotychczasowe wybory muzyczne i, niewątpliwie, zazdrośnie skrywaną bazę danych na mój temat, jest on w stanie generować długą listę muzyki wybraną specjalnie dla mnie. Istnieje też jednak tryb „muzyki podobnej do” – po wskazaniu na konkretnego artystę dobierane są utwory podobne. Wielu użytkowników narzeka, że jest to muzycznie mało inspirujące: nie zdajemy sobie z tego na co dzień sprawy, ale powstaje *naprawdę* bardzo dużo *naprawdę* bardzo podobnej muzyki. Wróćmy myślą do tych wszystkich upiornie szczegółowych klasyfikacji, o których wspomniałem na początku tego rozdziału. Choć oddani fani pewnie by mnie za tę opinię rozszarpali, to muszę nieśmiało przyznać, że piętnaście zespołów grających „industrialny psytrance” albo „atmosferyczny pogański black metal” tworzy... naprawdę podobne do siebie utwory. Ryzykując nocny atak ze strony rozjuszonych wielbicieli Swaróżyca, powiem nawet więcej: algorytm pokroju Aivy, sprzęgnięty z jakimś w miarę przytomnym generatorem tekstów, mógłby już dziś wygenerować coś, co przemknęłoby niezauważone o pierwszej w nocy na niejednej imprezie.

To powinno nas zaś uczulić na aspekt rynkowy rozwoju AI. Generatory tekstu, obrazu, filmu, muzyki i (być może już niedługo) narracji nie zawsze będą miały za zadanie wygenerowanie genialnego arcydzieła. Pierwszym zadaniem dla nich będzie raczej stworzenie krótkich tekstów dziennikarskich podsumowujących, kto strzelił w której minucie gola, funkcjonalnego wzoru tapety do biurowca albo 30-sekundowego klipu muzycznego towarzyszącego scenie wybuchającego wieżowca w najnowszym hollywoodzkim filmie akcji. Pozostawię waszej domyślności, które z tych trzech zadań już dziś naprawdę realizują algorytmy. A może wszystkie?

To zaś z kolei prowadzi nas nieco głębiej – ku refleksji na temat naszej własnej inteligencji oraz naszych własnych potrzeb. Jako „twórcy” względnie rzadko chwytamy Muzy za kostki, wydzierając Niebiosom okruchy transcendentnego Piękna. Dużo częściej

tworzymy funkcjonalne produkty, spełniające wymogi projektowe, czasem realizując przy tym wręcz wstydliwie proste algorytmy. Jako odbiorcy z kolei nie zawsze potrzebujemy muzyki o głębi strukturalnej *Wariacji Goldbergowskich*, aluzyjnej prozy budzącej egzystencjalny dreszcz oraz filmu skomponowanego jak przez malarza. Czasem potrzebujemy po prostu godziny relaksującej muzyki albo prostej gierki, w której będziemy mieli duży karabin, zastępy względnie prostych do rozsiekania potworów i nieco trudniejszego do pokonania bossa, z którego wypadnie jeszcze większy karabin.

Myślę, że najbliższe lata pokażą, że o ile spełnienie abstrakcyjnych, źle zdefiniowanych celów, jak „prawdziwe myślenie” albo „sztuczna ogólna inteligencja”, może nie nastąpić nigdy, to algorytmy zdolne do generowania tekstu, obrazu, filmu i muzyki będą w szybko rosnącym stopniu „zaludniać” naszą codzienność. Nawet jeśli jakość tworzonego przez nie produktu nieprędko dorówna temu, co jest w stanie zrobić dobry ludzki twórca, algorytmy takie będą miały jedną gigantyczną zaletę: będą w stanie cierpliwie, bez gadania, a być może nawet za darmo generować tysiące i miliony megabajtów treści, aby każdy mógł ujrzeć na ekranie komputera swoją okolicę, imiona i twarze swoich znajomych i poznać się bliżej ze swoim ulubionym celebrytą – albo wspólnie z nim rozwalić budynek zajęty przez terrorystów o twarzach dziwnie przypominającej grono pedagogiczne z liceum. Jest bowiem cała klasa produktów, których wykonanie wymaga tylko minimalnej kreatywności i artyzmu, za to kompletnego braku skrupułów estetycznych i moralnych. Te zaś programuje się dość opornie.

Twoje dane

W tym rozdziale pomówimy trochę o wizji, o której mówi się ostatnio tyle, że już niemal nam spowszedniała. Mowa o wizji globalnej sieci informacyjnej (to już jest), oplatającej każdego człowieka na planecie (już wkrótce), pobierającej i analizującej każdy aspekt naszego życia (i to niebawem). Wizja ta ma swoje jasne strony – jak choćby równy, demokratyczny dostęp do informacji i wiedzy dla każdego człowieka – oraz ciemne – pośród których najgroźniejszą jest chyba niebezpieczeństwo całkowitego zaniknięcia życia prywatnego i przekazania gigantycznej władzy nad nami tym, którzy mają dostęp do naszych danych.

Informacja to władza. W rozdziale II.1 wyobrażaliśmy sobie, jak to będzie żyć w świecie, w którym każdy człowiek posiadać będzie dostęp do danych pozwalających mu na skonstruowanie turbiny wiatrowej albo wyprodukowanie porcji insuliny. Mówiliśmy też nieco o ruchu na rzecz **otwartego dostępu do wiedzy** (*open source* i *open access*). Otwarcie dla wszystkich zainteresowanych dostępu do nauki, technologii i medycyny to specyficzna forma „rozproszenia władzy”.

W tym rozdziale powiemy trochę więcej o drugiej stronie tej samej monety: o otwarciu się dostępu do naszych danych i danych o nas. Ponieważ niniejsza książka ma być przede wszystkim o technologii, napiszę względnie niewiele o aspekcie prawnym i moralnym tego procesu. Bardzo mi natomiast zależy, żebyśmy przyjrzeni się bliżej stronie technologicznej i „sprzętowej” globalnej sieci teleinformatycznej, która pośredniczy w tym procesie. Czym właściwie jest dziś Internet? Ile mieści się w nim danych, jakimi ścieżkami one przepływają i gdzie tak naprawdę one się znajdują? Czym jest w praktyce owa osławiona *chmura danych*? Jakie dane generuje nasz smartfon, jakie dane zbierają nasi usługodawcy i co można na ich podstawie ustalić?

Jeżeli rzeczywiście boisz się Globalnej Maszyny, to warto poznać ją trochę bliżej. Spróbujmy zajrzeć poza ekrany naszych

urządzeń i pogmerać trochę w plątaninie obwodów i kabli, które kryją się za nimi. Bo ostatecznie, wierzcie lub nie, to wciąż jest po prostu wielka plątanina kabli.

Idźmy może jednak po kolei.

III.1. Anatomia globalnej maszyny

Mój pierwszy kontakt z komputerami miał miejsce na początku lat 90., więc choć nie uczestniczyłem w pionierskich latach komputeryzacji, po części „załapałem się” na pierwsze lata Internetu – choć początkowo tylko jako zafascynowany dzieciak. Mniej więcej w połowie szkoły podstawowej po raz pierwszy załogowałem się „do Sieci” – musiało to być przed 1998 rokiem, ponieważ na pewno nie było wtedy jeszcze Googla; prawdopodobnie w okolicach roku 1995. Kiedy z grupką innych dzieciaków staliśmy przed gigantycznym kineskopem szkolnego komputera, próbując zrozumieć, z czym właściwie mamy do czynienia, naszym głównym punktem wejścia był katalog stron internetowych Yahoo!, a wielkim odkryciem stała się AltaVista, do której „można było wpisać cokolwiek”. Niemal magiczną moc wydawał się mieć Ask Jeeves, czyli wyszukiwarka internetowa, której można było zadać pytanie. Znacznie wyraźniej pamiętam już czas, kiedy powstała Wikipedia, a polskiej edycji długo brakowało nawet elementarnych haseł – które zacząłem namiętnie dodawać i edytować w pierwszych kilku latach XXI wieku, już jako licealista.

Piszę o tym nie tylko po to, aby przespacerować się aleją dziadowych wspomnień, ale żeby wyjaśnić, jak ludzie „tacy jak ja” widzą Internet. „Sieć” od zawsze była dla mnie po prostu zbiorem połączonych ze sobą komputerów, a „surfowanie po internecie” oznaczało łączenie się z tymi komputerami, aby ściągnąć z nich dane lub poprosić je o wykonanie dla mnie jakiejś „usługi”. I... tyle. Pod koniec lat 90., kiedy wciąż dość umiarkowanie przestrzegano internetowych „zasad BHP”, względnie często dało się przejść od przeglądania czyjejś strony internetowej do przeglądania zawartości udostępnionych przez niego folderów – co zawsze bardzo lubiłem, bo czułem się wtedy jak „haker”. W tych momentach przeglądarka internetowa stawała się po prostu menedżerem plików – na cudzym komputerze! Do dziś pamiętam dreszcz ekscytacji, kiedy ze strony internetowej jakiejś amerykańskiej biblioteki dotarłem do macierzystego katalogu zawierającego tajemnicze foldery rozmaitych instytucji miejskich i uniwersyteckich – ktoś po prostu nie

dopilnował, aby odpowiednio ograniczyć uprawnienia gości. Krótko mówiąc, Internet zwykle się traktować jako sposób na łączenie ze sobą czyichś komputerów.

Dziś Internet jako taki stał się całkowicie „przezroczysty” i nie myślimy już w ogóle o jego stronie sprzętowej – przynajmniej nie bardziej niż myślimy o transformatorach, rozdzielniach i liniach przesyłowych wysokiego napięcia, gdy włączamy toster. Trudno o lepszą demonstrację tego faktu niż orzeczenie Rady Języka Polskiego, która już w 2002 roku zgodziła się ostatecznie na pisownię słowa „internet” również małą literą, argumentując, że w zastosowaniu codziennym nie używamy go jako technicznej nazwy na sieć komputerową, lecz pospolite określenie na jedno z mediów, jak telewizja czy radio.

Nie zmienia to jednak faktu, że jego zasadnicza struktura tak naprawdę nie uległa zmianie. Kiedy na ekranie mojego smartfona wyświetla się strona główna Facebooka, „komputer” tkwiący w jego chińskiej obudowie łączy się z serwerami Facebooka. Jeszcze niedawno niemal na pewno byłby to jeden z serwerów amerykańskich, a dane „biegłyby” którymś z kabli podmorskich po dnie Oceanu Atlantyckiego (od jakiegoś czasu część danych Europejczyków ląduje w jednym z czterech centrów danych na naszym kontynencie, skąd trafia do nas wiązkami światłowodów przecinającym kontynent wzdłuż i wszerz).

Jeżeli chcesz poczuć się jak „haker z filmów”, to otwórz na swoim Windowsie^[36*] aplikację o nazwie Wiersz poleceń (Command Prompt)^[37*], a następnie wykonaj polecenie „tracert www.facebook.com”. Polecenie to ujawni wszystkie etapy pokonywane przez Twoje dane na drodze do najbliższego serwera Facebooka. Ja przed chwilą zostałem połączony, siedmioma kolejnymi krokami, z adresem IP o numerze 31.13.81.36, co oznacza, jak sprawdziłem, serwery Facebooka na przedmieściach Dublina, gdzie mieści się zajmujące łącznie ponad 1 km^[2] centrum danych Facebook Clonee. W kolosalnych hangarach mieszczą się tam szafy wypełnione dyskami twardymi, połączone światłowodami o łącznej długości liczonej w setkach tysięcy kilometrów. Gdy więc wrzucam „na fejsa” najnowsze zdjęcie z imprezy, najpierw wędruje ono chwilę bezprzewodowo do mojego operatora komórkowego, lecz następnie nieprzerwanym ciągiem kabli do miejsca typu Centrum Danych Facebook Clonee, gdzie trafia na któryś ze zlokalizowanych tam dysków twardych. Warto to podkreślić: słynna „**chmura**”, w której przechowuje się dziś powszechnie dane, to w praktyce po prostu inna nazwa na globalną sieć tego typu centrów

danych^[38*]. To tam ostatecznie trafiają więc wszystkie wykonane przeze mnie na imprezie zdjęcia, jeśli tylko nie uprę się, aby, jak jaskiniowiec, przechowywać je u siebie lokalnie w telefonie. Jeżeli więc chcemy zrozumieć, co tak naprawdę dzieje się z naszymi danymi, warto zapoznać się bliżej z centrami danych.

Najpierw jednak jeszcze jeden krok. W rozdziale tym będziemy sporo mówić o danych i o jednostkach służących do pomiaru ich ilości. Prawdopodobnie większość z was jest już wstępnie zaprzyjaźniona z megabajtami i gigabajtami, jednak dalsze rzędy wielkości – konieczne przy mówieniu o internecie – są mniej znane. Zacznijmy więc od szybkiej orientacji w skali wielkości^[39*]:



Podstawową jednostką informacji jest bit (b), ale gdy mówimy o danych, często podstawą obliczeń staje się 8 takich bitów tworzących łącznie bajt (B). Przedrostki są standardowe: kilo-, mega-, giga-, tera-, peta-, eksa-, zetta- i jotta-, a każdy kolejny z nich oznacza liczbę tysięcy razy większą^[40*]. Warto poćwiczyć „skakanie” po powyższej skali wielkości: tysiąc gigabajtów to terabajt, tysiąc terabajtów to petabajt. Milion to przeskok o dwa ząbki: milion terabajtów to eksabajt. Teraz, mam nadzieję, będzie nam trochę wygodniej poruszać się w świecie danych – choć ich ilość jest dziś tak duża, że w praktyce i tak nie do wyobrażenia.



Centra danych

„Centrum danych” to nic innego, jak określenie na miejsce, w którym przechowuje się dużo danych. Jeżeli firma, w której pracujesz, gromadzi dużo danych, prawdopodobnie ma w swojej serwerowni centrum danych. Ponieważ pojęcie to nie

ma sztynwej definicji, trudno teŝ powiedzieć, ile ich jest. Portal Statista podaje liczbę 7,1 mln w 2021 roku. Co ciekawe, choć liczba danych na świecie szybko rośnie, liczba centrów danych maleje – małe, lokalne centra są zastępowane przez wielkie, zarządzane przez dużych graczy rynkowych.

Firmy typu Google, Amazon, IBM czy Microsoft mają dziesiątki gigantycznych centrów danych i corocznie otwierają kolejne. W momencie pisania tej książki Google oficjalnie chwalił się posiadaniem 23 centrów danych rozsianych po całym świecie^[41*], Amazon – 38, IBM – 60. Choć Facebook prawdopodobnie zbiera – łącznie w 15 centrach danych – ilościowo mniej danych niż wspomniani giganci, to jednak dane te są szczególnie wrażliwe, więc firmę tę wymienia się zwykle jednym tchem z głównymi graczami na rynku.

Istnieją ponadto rządowe centra danych, jak choćby gigantyczne amerykańskie Utah Data Center, które okryło się złą sławą w związku z ujawnionym w 2013 roku przez Edwarda Snowdena masowym gromadzeniem danych o zwykłych obywatelach przez amerykańską agencję rządową NSA. Zarówno państwowe, jak i prywatne centra przechowują trudną do oszacowania ilość danych, a uzyskanie na ich temat wiarygodnych informacji graniczy z cudem. Przykładowo, gdy mowa o pojemności Utah Data Center, wciąż cytowany jest raport „Forbesa” z 2013 roku, w którym oszacowano, że wynosi ona 3–12 eksabajtów (EB), jednak jego autor przyznał wówczas otwarcie, że w rzeczywistości liczba ta może być znacznie większa i prawdopodobnie będzie stale rosła.

Chińczycy są jeszcze bardziej tajemniczy w tym temacie. Chińscy giganci technologiczni jak Aliyun (Alibaba Group) i Tencent nie informują nigdy wprost, ile mają centrów danych, lecz chwala się liczbą „stref dostępności” na całym świecie – te utożsamia się czasem z regionami otaczającymi poszczególne centra danych, co jednak nie musi być prawdą. Aliyun ma ich 69; Tencent – 62. Według raportu singapurskiego banku DBS z 2018 roku^[65] w Chinach znajdowało się 1641 centrów danych, z tego 16 „ultrawielkich”. W żadnym miejscu nie są jednak wymienione konkretne pojemności, a rozmiary centrów wyrażane są w liczbie „szaf” (*cabinets*) – w których jednak może mieścić się nieznana liczba dysków o nieznaney pojemności^[42*]. Centra z kategorii „ultra-large” to te, które mają ponad 10 tysięcy „szaf”. W momencie publikowania tego raportu w budowie było 40 kolejnych takich placówek.

Tak duże jest zainteresowanie tematem i tak gęsta spowija go mgła tajemnicy, że niektórzy analitycy próbują zgadywać, ile danych jest w stanie przechowywać konkretne centrum, kalkulując „wstecz” na podstawie ilości konsumowanej przez nie energii. Zasilanie – i chłodzenie! – tysięcy metrów kwadratowych wypełnionych kontenerami z dyskami twardymi, to niebagatelne wyzwanie. W 2018 roku w innym raporcie „Forbesa”^[66] oszacowano, że wszystkie centra danych razem wzięte pochłonęły w 2016 roku 416 TWh – co oznaczałoby niemal 2% całej wyprodukowanej na świecie energii elektrycznej. Gdy ogląda się zdjęcia satelitarne centrów danych, często można dostrzec przyległe do nich stacje elektroenergetyczne wysokiego napięcia. Przy wspomnianym wyżej centrum danych Facebook Clonee znajduje się gigantyczna stacja 220 kV – jedyna stacja tak wysokiego napięcia zbudowana ze środków prywatnych na terenie Irlandii. Powstała wyłącznie z myślą o tym jednym centrum danych. Facebook inwestuje teŝ w farmy wiatraków, aby częściowo uniezależnić się od irlandzkiej sieci energetycznej.

Na podstawie tego typu informacji można próbować wywnioskować, ile danych mieszczą poszczególne placówki, jednak tylko w najgrubszym przybliżeniu – zwykle mówi się, że typowa zdolność pojedynczego dużego centrum danych to kilka lub kilkanaście EB. Portal Statista podawał w 2018 roku, że wszystkie centra danych na świecie mieszczą łącznie 547 EB, a TechRepublic szacuje^[67], że w 2021 roku jest to 1,3

ZB. Wszystkie te liczby z *grubsza* mają sens, tj. liczba ultradużych centrów danych, zarówno prywatnych, jak i państwowych, oraz ich objętość sumują się do skali mniej więcej zettabajta. Z drugiej strony do bardzo podobnych liczb prowadzą oszacowania wychodzące od zdolności produkcyjnych największych producentów dysków twardych na świecie. Zupełnie innym zagadnieniem jest oczywiście „ciemny”, tj. niejawni sektor przechowywania danych.

Na koniec ciekawostka technologiczna. Ze względu na dynamiczny rozwój Sieci, już dawno temu pojawiło się zapotrzebowanie na wysoką mobilność infrastruktury sieciowej. Krokiem w tym kierunku są **modularne centra danych**. Przykładowo, Google już w 2005 roku zaczął pracować nad upakowaniem zestawu „zasilanie + chłodzenie + dyski” w standardowym kontenerze transportowym: takim samym „blaszaku”, jaki przewozi się ciężarówkami, pociągami i kontenerowcami. Patent uzyskali w 2007 roku. Modularnymi centrami danych chwala się również Dell, HP, IBM, Microsoft i Oracle; pozostali gracze na „rynku danych” pewnie też je mają. Dzięki budowie modularnej centra danych można szybko i dyskretnie składać, rozbudowywać, demontować i przewozić, choćby i na drugi koniec świata.

Wspomniałem już wyżej, że połączenia bezprzewodowe odgrywają wiodącą rolę tylko na „czubkach” sieci telekomunikacyjnej. Użytkownicy telefonów komórkowych łączą się z najbliższą stacją bazową drogą powietrzną, jednak same stacje bazowe, w miarę konieczności obsługiwanie coraz większej liczby użytkowników, coraz częściej przekazują sygnał sobie nawzajem światłowodami. Użytkownicy internetu na komputerach stacjonarnych albo bezpośrednio wpinają się do sieci światłowodowej, albo połączenie takie posiada router bezprzewodowy („Wi-Fi”), z którym się łączą. Krótko mówiąc, bez względu na to, z jakiego korzystamy urządzenia, dzieli nas tylko kilka kroków od globalnej sieci światłowodowej.

Ciekawą alternatywą są połączenia satelitarne, „omijające” w ogóle sieć naziemną. Choć w ostatnich latach jest o nich coraz głośniejsze, dziś i w najbliższej przyszłości odpowiadać będą za niewielki ułamek ruchu sieciowego. Ich główną zaletą jest dostępność. Sieć Starlink, w momencie pisania tych słów obejmująca 1635 satelitów na niskiej orbicie, jest teoretycznie w stanie dostarczyć dane w każdy punkt globu. Różnica w kosztach i możliwościach przesyłu jest jednak gigantyczna i póki co głównym sposobem globalnej wymiany informacji jest światłowód^[43*]. Poznajmy się więc z nim nieco bliżej.

W skali lokalnej jest to pajęczyna kabli oplatających dziś każde większe miasto i docierających do coraz większej liczby miasteczek i wsi. Orange Polska, posiadacz największej infrastruktury światłowodowej w naszym kraju, chwalił się w 2017 roku, że dysponuje siecią o długości 100 tys. km. W skali regionalnej (czyli w Polsce wojewódzkiej) połączenia światłowodowe wyglądają na

mapach jak drogi wojewódzkie i krajowe, często omijające duże skupiska miejskie „obwodnicami”. W skali krajowej i międzynarodowej wyłania się zaś tzw. sieć szkieletowa (*backbone*) – główne „pęczki” światłowodów, przez które przesyłana jest zdecydowana większość danych w toku zwykłego użytkowania internetu. Ponieważ jednak internet jest tworem globalnym, dane muszą prędzej czy później „zejść pod wodę”. Tu wkraczamy w fascynujący świat podmorskich kabli telekomunikacyjnych, dzięki którym tak naprawdę istnieje *globalna* sieć wymiany informacji.



Kable podmorskie

Historia kabli podmorskich sięga roku 1865, kiedy to pierwszy funkcjonalny kabel telegraficzny między Europą i Ameryką Północną położył 211-metrowej długości parowiec SS Great Eastern, wówczas największy statek w dziejach żeglugi. Już sześć lat później, w 1871 roku (jakże potężny jest nasz pęd ku łączeniu się!) kable takie łączyły wszystkie zamieszkałe kontynenty i od tego czasu świat pozostaje skomunikowany z prędkością światła. Dziś szacuje się, że na dnie mórz i oceanów świata leży 426 aktywnych kabli światłowodowych – a w każdym razie taką konkretną liczbę podaje strona SubmarineCableMap.com, sponsorowana przez Huawei Marine Networks i uważana powszechnie za rzetelne źródło wiedzy.



Zasadnicza „magia” kabli podmorskich kryje się w wiązce cieniutkich jak włos, przezroczystych włókienek przewodzących światło. Gdy zaświeci się w jeden koniec takiego włoska, światło to wędruje niemal bez strat na drugi koniec^[44*] – informacja przekazywana jest więc za pośrednictwem „mruknięć światła”, a nie sygnału „wyrzeźbionego” w prądzie elektrycznym. Włókienka te otoczone są dla bezpieczeństwa gumą, plastikiem i stalowymi drutami, jednak łączna grubość typowego kabla to mało imponujące 2,5 centymetra. Na lądzie i w strefach przybrzeżnych kable takie zakopane są w ziemi, ale w większych odległościach od linii brzegowej leżą one po prostu swobodnie na dnie oceanicznym.

Niektóre z nich są względnie krótkie, jak choćby kabel Denmark-Poland 2 o długości 110 km, oddany do użytku już w 1991 roku, łączący Mieleno z Bornholmem. Jest to jeden z dwóch trafiających do naszego kraju kabli podmorskich. Inne to prawdziwe monstra, jak choćby TPE (Trans-Pacific Express) – łączący Stany Zjednoczone z Chinami, Tajwanem, Japonią i Koreą Południową, o łącznej długości 17 968 km – albo AAE-1 (Asia Africa Europe-1), biegnący Kanałem Sueskim, o łącznej długości około 25 000 km. Typowa przepustowość współczesnego kabla światłowodowego liczona jest w dziesiątkach lub setkach terabitów na sekundę (Tbps).

Tradycyjnie, najdłuższe i najbardziej „pojemne” kable podmorskie należały do dużych firm telekomunikacyjnych, jak GTT, Lumen, Vodafone, Verizon, China Telecom, Orange, Tata Communications czy Telecom Italia. W ostatnich latach mnożą się jednak kable finansowane przez gigantów internetowych. Spośród 13 kabli podmorskich^[45*] łączących obecnie Europę z Ameryką Północną dwa należą w całości do Google: kabel Dunant uruchomiony w styczniu 2021 roku (o przepustowości 250 Tbps) i kabel Grace Hopper (około 350 Tbps) planowany na 2022 rok, a w trzecim kablu (Havfrue, 108 Tbps) Google ma udział. Facebook nie ma „swoich” kabli transatlantyckich, ale pozostaje współwłaścicielem trzech (Havfrue, Amitie i MAREA). Podobna jest sytuacja na Pacyfiku. Ogółem, cztery firmy z branży komputerowej zdążyły dotychczas zainwestować w transoceaniczne kable podmorskie w stopniu, który daje im status właściciela lub współwłaściciela: Google, Facebook, Microsoft i Amazon.

Na Pacyfiku wykształcił się też ciekawy sojusz Google-Facebook: firmy te są wyłącznymi współwłaścicielami kabla PLCN (Pacific Light Cable Network), łączącego USA z Filipinami i Tajwanem o przepustowości 144 Tbps. W 2023 roku ma zaś zostać oddany kabel Echo (również 144 Tbps), łączący kontynentalne USA z Palau, Guam, Indonezją i Singapurem, sfinansowany po połowie przez Echo Cable Holdings i GU

Holdings. Nie trzeba jednak wielkich umiejętności detektywistycznych, by sprawdzić, że ta pierwsza firma należy do Facebooka, a ta druga do Googla.

Choć na co dzień śledzimy raczej postępy w rozwoju sieci bezprzewodowej (choćby słynny dziś przypadek sieci piątej generacji, czyli 5G), z roku na rok rozwijają się też po cichu rozmaite technologie światłowodowe. Bądź co bądź, bez względu na to, jak zaawansowany technologicznie jest mój telefon i jak potężna oraz sprytna jest najbliższa mi antena bazowa sieci komórkowej, jeżeli oglądany właśnie przeze mnie na telefonie film przechowywany jest na dysku znajdującym się – wróćmy do tego przykładu – na irlandzkich serwerach Facebooka, to i tak trafia do mnie kablami.

Bez względu na to, jak bardzo ukryty pod płaszczykiem współczesnej technologii byłby ten fakt, w istocie od strony czysto sprzętowej Internet to więc wciąż ta sama sieć spięta kablami. Wszystkie dane – a więc od strony treści „cały Internet” – ostatecznie znajdują się zaś po prostu na czyichś dyskach twardych^[46*]. Jak duży jest tak naprawdę „cały Internet”?

III.2. Ile jest danych?

Nikt nie wie z pewnością, ile jest danych zgromadzonych we wszystkich komputerach świata i jakich rozmiarów jest internet. Albo, mówiąc inaczej, uzyskanie wiarygodnych danych na ten temat graniczy z niemożliwością. Zacznijmy od szybkiej orientacji w skali wielkości. Same przedrostki omówiłem wyżej, ale teraz spróbujemy przetłumaczyć te wielkości na jakieś konkretne przykłady z życia codziennego.

Przy oszczędnym zapisie 1 MB powinien spokojnie wystarczyć do zapisania treści dowolnej książki. Oznacza to, że 40 milionów książek zgromadzonych w Bibliotece Kongresu zajmowałoby łącznie około 40 TB (terabajtów). Choć dyski o tej objętości są rzadkie, ekonomicznym sposobem przechowania tych danych byłyby na przykład 4 dyski 10-terabajtowe, łączny wydatek kilku tysięcy złotych.

Tekst to jednak bardzo oszczędny typ danych – dużo więcej przestrzeni dyskowej wymagają obrazy, dźwięki i filmy, zwłaszcza ze względu na galopujący wzrost ich jakości. Proste zdjęcia wykonywane starym telefonem „ważyły” kiedyś po kilkadziesiąt kilobajtów; fotki

wykonane nowoczesnym smartfonem lubią dziś zajmować po kilka megabajtów, a nawet więcej. Dłuższe filmy nawet średniej jakości zajmują już dziesiątki i setki megabajtów. Gdy wykonujemy zdjęcia i filmy w wysokiej rozdzielczości i zależy nam na jakości ich zapisu, szybko wchodzimy w strefę gigabajtów. Współczesne duże, złożone gry komputerowe miewają już po kilkadziesiąt GB^[47*]. Nic więc dziwnego, że dzisiejszy użytkownik urządzeń mobilnych potrzebuje dziesiątków i setek gigabajtów pamięci, a współczesne komputery domowe już zwyczajowo posiadają dyski terabajtowe.

Firma Seagate, jeden z największych na świecie producentów dysków twardech, podaje, że w 2021 roku średnia objętość sprzedawanych przez nich nowych dysków wynosiła 5,1 TB (jeszcze w 2015 roku był to 1 TB). W innych źródłach można z kolei wyczytać, że rocznie globalnie sprzedaje się około 500 mln dysków twardech, co dawałoby łączną objętość 2,5 ZB. Do problemu można też podejść inaczej. Szacuje się, że na świecie są około 2 mld komputerów osobistych. Nie każdy z nich jest supernowoczesny, a więc przy ostrożnym oszacowaniu, że ich dyski twarde mają średnią objętość 0,5 TB, wychodzi łączna objętość 1 ZB. Dyski te oczywiście nie zawsze pękają w szwach^[48*] – nawet więc po doliczeniu smartfonów^[49*] można bezpiecznie oszacować, że ilość „osobistych” danych zwykłych użytkowników urządzeń elektronicznych mieści się w okolicach 1–10 ZB.

Dużo trudniej jest oszacować „rozmiar Internetu”, czyli, można by chyba powiedzieć, danych udostępnionych publicznie. Choć nie każdy jego zakątek będzie nas dzisiaj interesował, to chyba dobra okazja, żeby pomówić nieco o „mapie Internetu”.



Deep web, dark web, Tor

Pojęcie deep web, czyli „głęboka sieć”, bywa źle rozumiane i mylone z dark web („ciemną siecią”) – oba pojęcia spowite są zaś zupełnie niepotrzebnie aurą tajemnicy i grozy.

Termin „deep web” zdefiniowany jest bardzo prosto: jako zbiór wszystkich treści, które są potencjalnie dostępne dla użytkowników internetu, ale do których nie prowadzą wprost wyszukiwarki internetowe typu Google, Bing czy DuckDuckGo. Aby zrozumieć ideę deep web, należy więc najpierw zrozumieć, jak działają owe wyszukiwarki. Te zaś opierają się na „pełzaczach” (*web crawler*), czyli programach komputerowych automatycznie odwiedzających strony internetowe i „indeksujące” je, czyli wciągające na listę stron „znanych” wyszukiwarce, a więc i podawanych użytkownikom w odpowiedzi na ich zapytanie (dziś zwykle robione są też przy okazji ich kopie). Podstawową metodą stosowaną przez crawlery jest podążanie za linkami. Prosty sposobem „ukrycia” czegoś przed wyszukiwarkami jest więc stworzenie strony internetowej, do której nie prowadzi żaden link. Czysto teoretycznie można by więc stworzyć gigantyczny serwis poświęcony, przypuśćmy, jeżozwierzom, zawierający artykuły, zdjęcia, filmy i inne dane, ale dopóki adres do niego podawany byłby „z ust do ust” i nie został opublikowany na żadnej innej stronie internetowej, wyszukiwarki nie wiedziałyby o jego istnieniu i nie kierowałyby do niego użytkowników wpisujących hasło „jeżozwierz”. Ów serwis należałby więc do deep webu.

Istnieją też treści, do których dostęp możliwy jest tylko po utworzeniu konta i zalogowaniu się. Prosty przykład: wszystkie artykuły prasowe, do których dostęp wymaga subskrypcji. Do deep webu należą więc również i moje własne teksty, które pisałem na przestrzeni lat do rozmaitych czasopism – jeśli tylko wydawcy prasowi umieścili je w internecie, ale wyłącznie dla płatnych subskrybentów. Nawet zaś jeśli dostęp do jakiejś strony jest bezpłatny, ale wymaga zalogowania się, proste crawlery „odbijają się” od strony logowania i nigdy nie dowiedzą, co kryje się po jej drugiej stronie (twórcy tych sprytniejszych crawlerów lubią bawić się w obchodzenie tego typu szlabanów). Autorzy danej strony mogą też zażyczyć sobie wprost, aby nie była ona indeksowana (w całości lub częściowo), na przykład umieszczając odpowiednią informację w specjalnym pliku „robots.txt” – wszystkie „przyzwoite” wyszukiwarki (w tym Google, Bing, Baidu i DuckDuckGo) respektują takie życzenie.

Krótko mówiąc, istnienie deep web zdefiniowane jest w odwołaniu do możliwości wyszukiwarek internetowych i wynika z ich niedoskonałości – oraz niechęci właścicieli stron do upubliczniania swoich treści. W pewnym sensie każda zwykła strona internetowa zaczyna w „sieci głębokiej”, a „wypływa na powierzchnię” (do „sieci powierzchniowej”, czyli „surface web” – tak czasami mówi się w kontekście deep web na „po prostu internet”) dopiero wtedy, gdy crawlery uzyskają do niej dostęp i umieszczają ją w swoich bazach.

Nikt tak naprawdę do końca nie wie, jakich rozmiarów jest deep web, choć często cytowane jest oszacowanie z 2001 roku^[69], że rozmiar „sieci głębokiej” około 500-krotnie przewyższa rozmiar sieci „zwykłej”, tj. zindeksowanej przez popularne wyszukiwarki. Z oszacowaniem tym trudno dyskutować (jak zmierzyć coś, czego z definicji nie da się systematycznie spisać?), choć ostrożniejsi komentatorzy podają często skromniejsze przypuszczenia. Wszyscy wydają się jednak zgadzać, że deep web znacznie przekracza rozmiarem „surface web”.

Dark web, czy też darknet to zupełnie inna koncepcja. Istotą „ciemnej sieci” jest celowe ukrywanie dostępu do pewnych treści oraz zwykle ponadto do danych identyfikujących autorów i użytkowników. O sieci typu darknet można też pomyśleć jako o „sieci w sieci”. W warstwie sprzętowej opiera się ona na tych samych urządzeniach i tym samym podstawowym „protokole komunikacyjnym” (czyli języku,

za pomocą którego komunikują się ze sobą maszyny podpięte do Internetu). Za sprawą odrębnego języka „wyższego poziomu”, którym maszyny mogą się ze sobą komunikować, a także osobnych aplikacji i osobnego systemu adresów wymiana danych zachodząca w „ciemnej sieci” dokonuje się jednak niejako *obok* zwykłej komunikacji.

Jeżeli rozumiemy termin „darknet” szeroko, to do kategorii tej zalicza się choćby BitTorrent – protokół służący do bezpośredniej wymiany plików pomiędzy użytkownikami, w 2021 roku obchodzący swoje dwudzieste urodziny. Choć czysto teoretycznie protokół taki może być wykorzystywany w dowolny sposób, to stanowi on w praktyce jeden z podstawowych sposobów przesyłu pirackiej muzyki, filmów, gier i oprogramowania. Szacuje się, że od kilku do nawet kilkudziesięciu procent danych – zależnie od czasu i miejsca – przesyłanych w internecie wędruje za pośrednictwem tego typu sieci wymiany plików.

Dzisiaj, kiedy mówimy o darknecie, zwykle mamy jednak na myśli bardzo określoną sieć tego typu: Tor. Od strony technicznej Tor opiera się na specyficznym sposobie przesyłania i szyfrowania danych, określanym jako „trasowanie cebulowe” (*onion routing*); Tor to w istocie skrót od słów „The onion router”. Jest to procedura wielokrotnego, warstwowego (stąd ta „cebula”) szyfrowania pakietów danych, która bardzo utrudnia identyfikację użytkowników. Przy pomocy przeglądarki obsługującej tego typu szyfrowanie, na przykład Tor Browser, można korzystać ze zwykłych zasobów internetu. Jest to nieco wolniejsze, ponieważ dane wędrują czasem przez kilka kontynentów, jednak znacząco zwiększa anonimowość użytkownika. To jedno z zastosowań Tor: możliwość swobodnego korzystania z internetu bez troszczenia się o to, że odwiedzane przez nas adresy są odnotowywane, analizowane i przypisywane do konkretnej osoby. Tor tradycyjnie cieszy się więc popularnością w krajach „na spektrum totalitaryzmu”.

Gdy za pomocą przeglądarki typu Tor Browser korzystamy ze zwykłych zasobów Sieci (na przykład wchodząc na stronę Wikipedii), nasze „zapytanie” adresowane jest do normalnych serwerów: w pewnym momencie „wyskakuje” więc z sieci Tor i trafia do „normalnej Sieci”. Istnieje też jednak cała alternatywna sieć „ukrytych serwerów”, do których dostęp możliwy jest wyłącznie poprzez przeglądarkę zgodną z Tor, a nie przy pomocy zwykłej przeglądarki internetowej. Adresy tych stron mają często postać na pozór losowych ciągów znaków i kończą się domeną o nazwie .onion (typowe domeny, z których korzystamy na co dzień, to .pl, .com, .edu, .gov i tak dalej). Gdy użytkownik łączy się z którąś z tego typu stron, komunikacja nigdy nie wychodzi poza sieć Tor, co zapewnia bardzo wysoki poziom anonimowości. Żadna technologia nie jest oczywiście doskonała i istnieją liczne znane słabe punkty Tor oraz długa historia udanych identyfikacji przestępców posługujących się tą siecią. Zwykły użytkownik, który nie jest przedmiotem zainteresowania organów ścigania, w praktyce korzysta jednak z Tor anonimowo.

Co znajduje się w darknecie? Najprościej przekonać się samemu. Zainstalowanie przeglądarki typu Tor Browser i znalezienie w „zwykłym” internecie linków do stron typu .onion nie nastręcza wielkich trudności. Ba, istnieją nawet „darknetowe wyszukiwarki”, indeksujące strony typu .onion. Ogólnie, nie trzeba dziś szczególnego wysiłku ani otrząskania informatycznego, aby poszperać sobie po darknecie^[50]. Przynajmniej po jego części – istnieje też bowiem „onionowy” odpowiednik deep webu: należą do niego na przykład sklepy („bazary”), na które trzeba otrzymać zaproszenie.

Jak łatwo się domyślić, głównym obszarem zainteresowania użytkowników „ciemnej sieci” jest wszystko, co nielegalne^[51]. Na jednym z popularnych katalogów stron .onion główne kategorie to: „usługi finansowe” (głównie kryptowaluty i fikcyjne karty kredytowe), „usługi komercyjne” (broń, paszporty, róg nosorożca, tania

elektronika), whistleblowing, hakowanie^[52*], pliki, erotica, narkotyki. Ponieważ bywalcy darknetu są bardzo dumni ze swoich powiązań z kontrkulturą, anarchistami, bojownikami o wolność i wszelkiej maści „antysystemowcami”, popularnym motywem są mniej lub bardziej pretensjonalne blogi i kolekcje materiałów o zabarwieniu „wywrotowym”: poradniki, jak uwikłać polityka w pornografię dziecięcą, skonstruować bombę z nawozów azotowych i sparaliżować strony rządowe atakami „hakerskimi”. Jak łatwo zgadnąć, zwykle trudno jest zorientować się, czy te wszystkie groźnie wyglądające strony robione są na serio, czy też ktoś realizuje na nich po prostu swoje stoletnie fantazje. Co więcej, zdarzyło mi się znaleźć nawet „przepis na bombę atomową”, rzekomo w stu procentach oparty na ściśle tajnych materiałach militarnych, a w rzeczywistości inspirowany chyba jednak głównie komiksami o superbohaterach i diagramami z Wikipedii.

Część tego „targowiska różności” jest oczywiście najzupełniej rzeczywista. Przykładowo, handel narkotykami przez darknet to realne zjawisko. W 2017 roku we wspólnym raporcie europejskiej agencji ds. narkotyków EMCDDA i Europolu^[71] oszacowano, że narkotyki odpowiadają za 62% całego rynku „darknetowego”.

„Sieć głęboka” i „ciemna” to dwa ciekawe zjawiska, więc gdyby książka niniejsza była poświęcona wyłącznie internetowi, na pewno pomówilibyśmy sobie o nich więcej. Mało tego, byłby to pewnie pretekst do bliższego poznania się z kryptowalutami typu BitCoin i technologią blockchain, na której się opierają^[53*]. Musimy się jednak skupić. Ile jest tego całego internetu?

W światku inżynierów i programistów trzymających rękę na pulsie Sieci przyjęło się dzielić jej anatomię na dwa zasadnicze segmenty: *core* („rdzeń”) i *edge* („krawędzie”)^[54*]. Do „rdzenia” wlicza się infrastrukturę sieciową dużych firm, kierujących znaczną ilością ruchu, oraz usługodawców oferujących masowo serwery i przestrzeń dyskową. „Krawędzie” Internetu to zaś urządzenia należące do małych firm i zwykłych „szarych” użytkowników. Czasem mówi się o nich również jako o „wierzchołkach” Sieci.

Co ciekawe, choć początkowo Internet „należał” głównie do użytkowników, z czasem rola „rdzenia” zaczęła rosnąć. International Data Corporation (IDC), organizacja monitorująca rozwój technologii cyfrowych, szacuje^[72], że w 2019 roku miał miejsce „punkt przegięcia” i pierwszy raz w historii Internetu większość danych w globalnej „datasferze” rezydowała w „rdzeniu”, a nie w jej „wierzchołkach”. Głównym czynnikiem za to odpowiedzialnym jest przechowywanie danych „w chmurze”: jak już wspomniałem wyżej, wbrew nazwie i wiążącym się z nią skojarzeniom, wysłanie danych „do chmury” oznacza de facto przesłanie ich do centrów danych firm typu Amazon, Microsoft, Google, Alibaba czy IBM^[55*]. Choć Internet rozpoczął jako

twór rozproszony, niemal idealnie demokratyczny, z każdym kolejnym rokiem coraz bardziej koncentruje się wokół kilku gigantów, oferujących przestrzeń dyskową i zarządzających infrastrukturą sieciową.

Zobaczymy, przykładowo, co oznacza w praktyce zgoda na to, aby wykonywane przeze mnie smartfonem zdjęcia były przechowywane „w chmurze” – coś, do czego użytkownik nowego telefonu jest dziś namawiany namolnie od pierwszych sekund jego użytkowania. Przy konwencjonalnym zapisie fotografia trafia po prostu do pamięci wewnętrznej telefonu, gdzie – plus minusz atak hakerski – pozostaje względnie prywatna. Przy zapisie „w chmurze” plik ten wędruje do najbliższego centrum danych należącego niemal na pewno do jednego z wymienionych wyżej gigantów, po drodze podlegając zresztą szczegółowej analizie (o czym niżej). To właśnie dlatego rola „wierzchołków” stopniowo maleje.

Jeśli wierzyć oszacowaniom IDC, w 2018 roku globalna „datasfera” miała objętość 33 ZB, a do 2025 roku ma urosnąć do 175 ZB. Wówczas zaledwie 30% danych będzie znajdowało się po stronie „wierzchołków” i – jeszcze ciekawsza statystyka – tylko 50% danych będzie tam *generowanych*.

Musicie mi wybaczyć jeszcze trochę statystyk, ale wydaje mi się, że dużo łatwiej nam będzie zrozumieć, co się dzieje dzisiaj z internetem i naszymi danymi, jeżeli zaprzyjaźnimy się z „cyframi”. Jedną z największych na świecie firm w branży technologii komunikacyjnych, Cisco, publikuje corocznie „raport o stanie internetu”. Co ciekawe, nie pada tam żadna konkretna liczba reprezentująca „ilość danych w internecie”. Autorzy raportu podchodzą do tematu nieco inaczej, próbując raczej odpowiedzieć na pytanie, ile danych corocznie „przepływa” przez internet. W najnowszej wersji raportu^[73] przedstawiono stan bieżący i prognozy do roku 2023. Parę wyimków:

- W 2023 roku liczba osób z dostępem do internetu osiągnie 5,3 miliarda (czyli 2/3 globalnej populacji).
- W 2021 roku globalny ruch sieciowy ma wynieść 3,3 ZB (tyle danych zostanie przesłanych przez internet).
- Około 82% przesyłanych do użytkowników treści to dane wideo.

Parę komentarzy. Po pierwsze, skala *ruchu* w internecie – czyli ilości przesyłanych „po kablach” danych – to coś zupełnie innego niż jego objętość. Prosty przykład. Aktualny rekordzista pośród najbardziej popularnych filmów na YouTubie, piosenka dla dzieci „Baby Shark

Dance”, zarejestrował 9,2 mld wyświetleń (tak, liczba jego odtworzeń przekracza liczbę ludności na świecie). Sam ów dwuminutowy film rezydujący gdzieś na serwerach Googla (od 2006 roku YouTube należy do Googla) może zajmować nie więcej niż 100 MB, jednak – w całości lub fragmentach – przemierzył internet miliardy razy.

Owe zetabajty danych kursujące dzień i noc po internecie to więc w przeważającej większości drobny ułamek najpopularniejszych filmów na YouTube, Netflixie i innych popularnych platformach streamingowych. Większość filmów, zdjęć i dokumentów wrzuconych do internetu nigdy nie zostanie obejrzana więcej niż kilka razy. W ciągu jednego dnia na Instagram trafia 67 milionów nowych zdjęć, a na Facebooka 300 milionów zdjęć i 1,1 miliarda (!) wypowiedzi tekstowych. W tym samym czasie na YouTube umieszczane są filmy trwające łącznie 82 lata^[56].

Ostatnia porcja statystyk. W cytowanym wyżej raporcie Cisco podano, że w 2023 roku będzie nieco ponad 5 miliardów użytkowników internetu. Co ciekawe, liczba *urządzeń* z dostępem do internetu osiągnie wówczas 29,3 miliarda, a najszybciej rosnącą grupą są te, z których ludzie nie korzystają bezpośrednio, a za pośrednictwem innych urządzeń: nadajniki GPS w samochodach, „sprytne” (smart) urządzenia domowe (kamery, domofony, odkurzacze), bransoletki medyczne, nadajniki/odbiorniki przemysłowe i tak dalej. W raporcie tym oszacowano, że mniej więcej połowa wszystkich *połączeń* będzie typu M2M (*machine-to-machine*), a odsetek ten będzie tylko rósł. W coraz większym stopniu globalna Sieć składa się więc z maszyn „gadających między sobą”.

Co więcej, to właśnie wymagania połączeń typu M2M wymuszają dziś rozwój strony sprzętowej Internetu, zwłaszcza ze względu na transfer danych ultraszybki i z minimalnymi opóźnieniami. Ludzie są niewyobrażalnie powolnymi istotami z punktu widzenia komputerów. Opóźnienie rzędu 100 milisekund (czyli 1/10 sekundy) jest dla człowieka praktycznie niezauważalne, a komputerowi musi dłużyć się w nieskończoność. Rozwój telekomunikacji w ostatnich dekadach to w dużym stopniu walka o zmniejszanie tzw. „latencji”, czyli opóźnienia w przekazie sygnału. Standard 5G uwzględnia tzw. ultraniską latencję, zbliżoną do 1 ms, a już teraz mówi się o kolejnej generacji, 6G, mającej między innymi sprowadzić latencję do skali mikrosekund. Jest to motywowane głównie wymaganiami połączeń M2M. Drugim aspektem rozwoju sieci telekomunikacyjnych, który ma związek raczej z rosnącą liczbą urządzeń, a nie potrzebami ludzkich użytkowników sieci, jest zdolność jednej anteny do obsługiwanie

coraz większej liczby urządzeń^[57*]. Pamiętacie roje autonomicznych dronów, o których pisałem w rozdziale I? To właśnie z myślą o obsłudze tego typu „rojów” urządzeń, które muszą błyskawicznie dzielić się gigantycznymi porcjami danych – z których większość nigdy nie trafia do człowieka, ponieważ służą one wyłącznie ich prawidłowemu funkcjonowaniu (co nie znaczy, że nie są gromadzone...) – projektuje się przyszłość sieci telekomunikacyjnych.

Obraz Sieci, który wyłania się z tych statystyk i tendencji rozwojowych, jest bardzo interesujący.

Większość danych tworzonych przez użytkowników trafia do wielkich centrów danych i już ich nie opuszcza. Lokalne przechowywanie danych przez użytkowników staje się coraz mniej istotne.

W coraz większym stopniu nie są to dane „świadomie upublicznione” (na przykład w formie zdjęcia, które ktoś postanowił opublikować). To raczej otaczające nas urządzenia generują coraz więcej danych w toku swojej pracy (o tym, jakie są to dane, powiem nieco więcej w następnym rozdziale).

Prędkość transferu danych stale rośnie, a opóźnienia występujące przy komunikowaniu się urządzeń maleją. Dąży się do tego, aby maszyny komunikujące się ze sobą za pośrednictwem Internetu w ogóle „nie czuły”, że komunikują się z inną maszyną, a sięganie do „własnej” pamięci powinno być równie łatwe i szybkie, co sięganie do pamięci dowolnego połączonego urządzenia^[58*].

Po tej szybkiej „lekcji anatomii Internetu” czas przyjrzeć się bliżej danym, które w nim krążą.

III.3. Człowiek i jego dane

Zacznijmy od jednego, bardzo ważnego pojęcia. Może tego nie wiecie, ale nie wszystkie dane są sobie równe. Są bowiem dane i są **metadane**.

Metadane to dosłownie „dane o danych”. Najprostszym przykładem metadanych jest swojski kartkowy katalog biblioteczny. Innym: adnotacja „Jasiek z dziećmi, Wisła 1992” wykonana długopisem na tylnej stronie zdjęcia. Nas interesować jednak będą metadane elektroniczne – dane elektroniczne o danych elektronicznych.

Jest jeden fundamentalny powód, dla którego powinniśmy być czujni w temacie „naszych” metadanych, być może nawet bardziej niż w przypadku „naszych” danych. Okazuje się bowiem, że o ile istnieje szereg bardzo rygorystycznych przepisów regulujących prawo dostępu do danych (jak choćby słynne RODO), to metadane stanowią coś w stylu... szarej strefy. Idźmy może jednak po kolei.



Metadane

Istnieje już dziś cała teoria metadanych. Ja proponuję myślenie o trzech „poziomach” tego zjawiska; nie wiem, czy odpowiada temu jakaś formalna klasyfikacja i czy nie gubię w tym wszystkim czegoś istotnego, ale trudno. To moja książka!

Zacznijmy od pierwszego poziomu: **metadane zamknięte w naszych plikach**.

Chociaż mogłoby się wydawać, że pliki komputerowe – tekstowe, graficzne, filmowe i wszelkie inne – to po prostu „pakiety danych”, w rzeczywistości precyzyjniej byłoby je określić jako „opisane pakiety danych”. Wróćmy do zwykłej papierowej fotografii i jej opisu wykonanego długopisem. Jeżeli chcemy zrozumieć metadane elektroniczne, wyobraźmy sobie, że gdy otrzymujemy lub wysyłamy gdzieś zwykłą fizyczną fotografię, to tak naprawdę jest to zawsze koperta zawierająca nie tylko samą fotkę, ale również formularz zawierający pewną liczbę pól, zgodnie z pewnym międzynarodowym standardem. Dane zawarte w owym formularzu to właśnie metadane.

Ot, wykonujemy zdjęcie smartfonem. Sama „treść” zdjęcia to tylko część powstałego w tej chwili w pamięci telefonu pliku, ponieważ automatycznie zostaje do niej dołączony pakiet informacji – metadanych – niemal na pewno w standardzie o nazwie EXIF. Typowy zestaw metadanych EXIF może zawierać: nazwę producenta i typ aparatu, datę i godzinę wykonania zdjęcia, współrzędne geograficzne i garść innych danych. Pliki dźwiękowe i filmowe bardzo lubią mieć „wypasione” standardy metadanych. Przykładowo, gdy mój smartfon zapisuje wykonane przeze mnie filmy – nawet durne trzysekundowe nagranie wykonane po przypadkowym dotknięciu przycisku nagrywania – przewiduje w ich opisie rubryki takie jak: „reżyser”, „producent” i „wydawca”, „kompozytor” i „dyrygent”, a nawet „okres historyczny”, „nastrój” i „gatunek muzyczny”. Sprawdziłem na chybił trafił parę swoich plików i aplikacje na moim telefonie nie wydają się korzystać z tych rubryk – a jednak pozostają one do dyspozycji.

Tak używało się terminu „metadane” pierwotnie i tak wciąż mogą zareagować na to słowo osoby „wykształcone informatycznie”. To jednak nie koniec. Drugi poziom to **generowane automatycznie metadane na temat naszej aktywności w sieci telekomunikacyjnej**.

Pomyślmy o rozmowie telefonicznej albo SMS-ie. Każda aktywność w sieci telekomunikacyjnej powoduje wygenerowanie zapisu na jej temat. W przypadku

połączenia głosowego zapis taki nosi nazwę CDR (Call Detail Record^[59*]) i zawiera m.in.: godzinę rozpoczęcia i czas trwania rozmowy, numery telefonów rozmówców, a także informację o tym, przez jaki punkt dostępowy rozmówcy połączyli się z siecią (na przykład z którą stacją bazową telefonii komórkowej). Bardzo podobne zapisy powstają przy każdym połączeniu się z Internetem: ten i ten użytkownik (identyfikowany poprzez tzw. numer IP) o tej i o tej godzinie połączył się z tym i tym adresem.

Na nieco wyższym poziomie sprawdzane i zapisywane są też nieco bardziej szczegółowe dane o naszych „zachowaniach”: nie tylko adresy stron, z którymi się łączę, ale też czas spędzony na ich oglądaniu, wysyłane przeze mnie zapytania, kliknięcia w reklamy, a w pewnym przypadku nawet ruchy myszką. W przypadku smartfonów – które zaczynają dominować jako główne narzędzie do korzystania z internetu (i już niedługo „ruszanie myszką” będzie pewnie brzmiało jak „strzelanie joystickiem”) – rolę „źniwiarza metadanych” przejmują aplikacje mobilne, którym w momencie instalacji w praktyce pozwalamy na wszystko.

To jednak nie koniec. Trzeci poziom to **metadane wygenerowane czynnie na podstawie naszych danych**.

I tutaj kryje się prawdziwa kopalnia złota. Niektóre (meta)dane tego typu zapewniamy sami, na przykład oznaczając znajomych na zdjęciach na Facebooku odpowiednimi „tagami”. Większość jest dziś jednak generowana automatycznie. Już w 2010 roku Facebook zaczął „podpowiadać” użytkownikom tego typu tagi – „karmiąc się” oczywiście danymi treningowymi zapewnianymi przez gorliwych użytkowników oznaczających wszystkich swoich znajomych na wszystkich swoich zdjęciach. Dziś nie jest to więc już system „podpowiadania tagów”, tylko po prostu automatyczna identyfikacja wszystkich osób występujących na wszystkich zdjęciach nadsyłanych do serwisu. Zdjęciu takiemu towarzyszy więc pakiet metadanych: jaki użytkownik, kiedy i skąd przysłał zdjęcie, kto się na nim znajduje i tak dalej. Pakiet ten nie jest już częścią pliku ze zdjęciem, jak metadane poziomu pierwszego, lecz osobnym zapisem w bazie danych. Wiemy też, że na podstawie naszej aktywności w internecie – co czytamy i oglądamy, z kim się zadajemy, jakie treści „polubiliśmy” – mogą być i są tworzone indywidualne profile psychologiczne, przypisane do naszego konta w danym serwisie. To też metadane, jeśli posługiwać się tym pojęciem w oryginalnym, ogólnym sensie.

Nikt tak naprawdę nie wie, jakie ilości tego typu metadanych są dziś generowane przez dysponentów naszych „danych pierwotnych”, którymi są nasi operatorzy telekomunikacyjni, producenci naszych telefonów i twórcy użytkowanych przez nas aplikacji i serwisów społecznościowych. Wiemy tylko, że wszystkim im żywo zależy na jak najpełniejszym „wyżęciu” naszych danych z wszelkich informacji, które można w nich odnaleźć. Rzecz w tym, że prywatne dane – jak fotografie czy filmy – są jednak w pewnym stopniu chronione. Metadane – praktycznie wcale.

W 2013 roku Edward Snowden ujawnił istnienie szeregu programów, nadzorowanych przez amerykańską agencję wywiadowczą NSA, mających na celu gromadzenie danych o użytkownikach sieci telekomunikacyjnych. Pośrodku globalnej burzy medialnej na temat masowej inwigilacji przedstawiciele rządu amerykańskiego uspokajali, że nie gromadzą żadnych „danych osobistych”. Ówczesny prezydent USA Barack Obama powiedział:

„Nikt nie podsłuchuje twoich rozmów”. Przypuśćmy, że tak rzeczywiście jest – że nikt nie gromadzi moich „danych”. Co jednak z metadanymi? W pewnych przypadkach nie tylko *wiemy*, że są gromadzone, ale wręcz istnieją prawa *nakazujące* ich gromadzenie, na przykład firmom telekomunikacyjnym, również w Polsce^[60*]. Otwarte jest pytanie, na ile dane te są w naszym kraju udostępniane instytucjom rządowym. W czerwcu 2013 roku ujawniony został dokument, z którego wynikało, że amerykańska firma telekomunikacyjna Verizon (ich odpowiednik Orange) ma obowiązek przekazywania agencji NSA zapisów CDR – co w praktyce oznacza pełną wiedzę na temat tego, kto z kim rozmawia, skąd i jak długo.

Ile można wywnioskować tylko na podstawie tego typu zapisów? W 2009 roku niemiecki polityk Partii Zielonych Malte Spitz postanowił dowiedzieć się, jakie dane na jego temat gromadzi Deutsche Telekom, który dostarczał mu usługi komórkowe. Rozpoczęła się żmudna batalia sądowa, ponieważ DT za żadne skarby świata nie chciał się tymi danymi podzielić; ostatecznie Spitz wygrał i otrzymał zestaw „swoich” metadanych za okres 6 miesięcy. Był to tak duży pakiet danych, że Spitz postanowił poprosić o pomoc w ich „rozczytaniu” dziennikarzy z „Zeit Online”, którzy za jego zgodą zamieścili w internecie interaktywną mapę, stworzoną wyłącznie na podstawie metadanych DT, pozwalającą każdemu zainteresowanemu śledzić 6 miesięcy z życia Malte Spitz. Ot, 1 września 2009 roku: Spitz przyjechał pociągiem do Monachium, gdzie o 2 po południu spędził nieco czasu w restauracji, najprawdopodobniej „Das Kloster” w pobliżu dworca Ostbahnhof. Wieczorem wrócił do Berlina samolotem – lądował o 11 wieczorem na lotnisku Tegel. Zaledwie kilka „punktów danych”, a jakże żywy i szczegółowy obraz czyjegoś życia. A to był rok 2009! Dziś przeciętny użytkownik smartfona w krajach wysoko rozwiniętych korzysta z niego przez kilka godzin dziennie i wykonuje na nim od kilkudziesięciu do kilkuset „działań” pozostawiających ślad w postaci metadanych^[61*], przy czym liczba ta błyskawicznie rośnie.

Żeby upewnić się, na czym polega moc metadanych, przeanalizujmy konkretny przypadek. Siedzisz ze znajomymi w mieszkaniu jednego z was. Ktoś wykonuje zdjęcie i udostępnia je na Facebooku. Fotografia ta błyskawicznie wędruje do któregoś z centrum danych Facebooka, gdzie ulega od razu „przemienieniu” do postaci użytecznych metadanych. Jakich, przykładowo?

O której godzinie i – jeżeli aparat zapisuje takie dane – pod jaką długością i szerokością geograficzną miało miejsce spotkanie

(informacje te zawierają metadane poziomu pierwszego wygenerowane przez smartfona w momencie robienia zdjęcia).

Jeszcze raz o której godzinie i – teraz już na pewno – w jakim mniej więcej miejscu miało ono miejsce (w metadanych poziomu drugiego wygenerowanych przez smartfona w momencie wysyłania zdjęcia znajduje się informacja o najbliższej stacji bazowej sieci komórkowej).

Kto się spotkał (tę informację zawierają z kolei metadane poziomu trzeciego wygenerowane przez Facebooka dzięki algorytmowi rozpoznawania twarzy). Na tym etapie mogą być generowane ponadto najrozmaitsze dodatkowe dane: jaki jest wyraz twarzy osób na zdjęciu, jakie są na nim przedmioty oraz jakie napisy i tak dalej. Dziennikarze śledczy wyłazą ze skóry, żeby dowiedzieć się, jakiego typu informacje „wyciągane” są tak naprawdę z naszych zdjęć, z kiepskimi rezultatami.

Możemy być więc pewni, że w momencie, gdy w jakikolwiek sposób wchodzimy w interakcję z naszym telefonem, pozostawia to po sobie trwały ślad, który zostaje scalony z naszym „profilem elektronicznym” i przeanalizowany. Każdy duży gracz na rynku teleinformatycznym prowadzi tego typu profil – stopień wymiany danych między nimi pozostaje już niejasny (zwłaszcza na linii firmy prywatne–rządy). Co istotne, z każdym kolejnym rokiem rozszerza się rozumienie tego, czym jest owa „interakcja”. Jest poniekąd oczywiste, że wysyłając SMS-a, zostawiam w bazach danych ślad po tym zdarzeniu – łatwo to uzasadnić choćby koniecznością naliczania nam opłat przez operatora telekomunikacyjnego. Zdążyliśmy się przyzwyczaić do tego, że zapamiętywana i analizowana jest też każda odwiedzana przez nas strona internetowa – choć tak naprawdę nie jest aż tak oczywiste, dlaczego dane te *muszą* być gromadzone. To jednak tylko wierzchołek góry lodowej. Dzisiaj nawet telefon, który po prostu nosimy ze sobą, nawet go nie dotykając, bezustannie nadaje... na nasz temat.



Co Google wie

W 2018 roku grupa programistów i inżynierów przeprowadziła ciekawe badanie^[77]. Nabyli kilka smartfonów wyposażonych w najpopularniejszy obecnie system operacyjny dla urządzeń mobilnych, Android, po czym całkowicie je „wyzierowali”, przywracając ich ustawienia fabryczne. Następnie zainstalowali na nich przeglądarkę Google Chrome oraz aplikację Google Play, służącą do pobierania innych aplikacji, tworząc specjalnie dla tego celu nowe konto Google. Plan był taki, aby przez 24 godziny śledzić, jakie informacje trafiają na serwery Googla w toku normalnego użytkownika smartfona – telefony zostały więc wyposażone w dodatkowe, ukryte narzędzie, pilnie monitorujące ruch sieciowy dokonujący się „w tle” zwykłego funkcjonowania telefonu, bez wiedzy użytkownika.

Pierwszy wynik: w ciągu 24 godzin telefon wysyłał informacje do któregoś z licznych serwerów Googla 2200 (słownie: *dwa tysiące dwieście*) razy – 90 razy na godzinę. Spośród tych 2 tysięcy „raportów” 450 stanowiło informację o położeniu użytkownika – ta była więc odnotowywana średnio co 3 minuty. Badanie Schmidta pokazało przy okazji potęgę analizy danych: do raportu o położeniu „dopisywany” był wydedukowany (na pewno na podstawie prędkości, ale być może również i samego położenia oraz czujników przyspieszenia, jeżeli takie znajdują się w telefonie) przypuszczalny sposób przemieszczania się: pieszo, na rowerze, samochodem, pociągiem^[62*]. Jednym z dużych źródeł informacji na temat położenia były sieci Wi-Fi dostępne w otoczeniu. Smartfony stale „skanują” otoczenie w poszukiwaniu dostępnych sieci; w ciągu jednego tylko 15-minutowego spaceru odnotowano wysłanie 9 raportów zawierających listę dostępnych dla urządzenia sieci. W raportach tych pojawiło się łącznie 100 identyfikatorów – a każdy z nich to niejako osobna informacja: „użytkownik o godzinie XX:YY:ZZ był w zasięgu sieci «Kawiarnia U Jędrusia» i sieci «Kasztanowa 13», ale przestał być w zasięgu sieci «Szkoła Podstawowa nr 3»”. Druga duża grupa informacji trafiała na serwery zajmujące się gromadzeniem informacji dla reklamodawców (usługi AdMob, AdSense, Google Ads i inne). Są to już najrozmaitsze informacje, zbierane głównie ze sposobem korzystania z treści dostępnych w internecie, zbieranych szczególnie obficie, gdy dana aplikacja albo strona internetowa korzysta z usług Googla. Choć czysto teoretycznie informacje te są „zanonimizowane”, w rzeczywistości wysyłane ze smartfona raporty zawierają identyfikatory pozwalające na połączenie ich z konkretnym użytkownikiem^[63*]. Łączna objętość danych wysłanych w ciągu doby to 11,6 MB; dla porównania: dokument zawierający całą treść niniejszej książki zajmuje mniej więcej 0,5 MB.

Jeżeli myślicie w tym momencie, że jesteście „bezpieczni” ze względu na swój sposób korzystania z telefonu, najprawdopodobniej się mylicie. Po pierwsze, w trakcie opisanego wyżej testu celowo unikano korzystania z najbardziej „oczyszczonych” usług Googla, jak Gmail albo YouTube. Jedynymi produktami tej firmy były aplikacje Google Play i Chrome. Ba, przeprowadzono dodatkową, zupełnie „odchudzoną” wersję eksperymentu, w trakcie której z aplikacji tych w *ogóle nie korzystano* – były one uruchomione wyłącznie „w tle” i nie zostały w ciągu 24 godzin użyte ani razu. W takim przypadku do Google trafiło „tylko” 900 paczek danych, z czego około 300 zawierało informację o położeniu. Skądinąd wiadomo, że smartfony gromadzą dziś informację o położeniu nawet wtedy, gdy wyłączone są w nich funkcje GPS i Wi-Fi oraz funkcja „Śledź moje położenie”^[78]. Źródłem informacji jest raczej sam system – Android – należący do „ekosystemu” Googla.

Po drugie, nawet jeżeli wydaje się wam, że nie uczestniczycie w ogóle w tym ekosystemie, będąc na przykład użytkownikami telefonów produkowanych przez Apple, w rzeczywistości i tak jesteście jego częścią. Autorzy opisywanego tu raportu przeprowadzili bliźniaczy eksperyment dla iPhone’a, wyposażonego w system operacyjny iOS, na którym zainstalowana była przeglądarka Safari. Wyniki nie napawają optymizmem. Przede wszystkim, Apple posiada swoją wersję wszystkich

usług Google, na przykład serwery gromadzące informację o położeniu użytkownika albo dane interesujące reklamodawców. Łącznie, podczas gdy typowy telefon z Androidem wysyłał raporty 90 razy na godzinę, typowy telefon z iOS-em – 70 razy, z czego 20 razy łączy się z serwerami Apple'a, a 50 razy z serwerami... Google. Nie zapominajmy, że miliony aplikacji i stron internetowych opierają się na usługach Google dla reklamodawców i twórców – nawet więc telefon należący do zupełnie innego „ekosystemu technologicznego”, jak choćby „uniwersum Apple'a”, tak naprawdę uczestniczy w świecie Google.

Zauważmy, że tego typu nieustannie płynący strumień danych nie składa się z treści, które *my generujemy*. Możemy w toku naszego codziennego korzystania z urządzeń elektronicznych nie napisać żadnego maila i nie wykonać żadnego zdjęcia, a i tak jesteśmy źródłem potężnej ilości danych na nasz temat. Co istotne, choć wszyscy zgadzają się, że mój email do kochanki albo zdjęcie niepotrzebnie wykonane o drugiej w nocy w pomrocznym stadium imprezy należą do sfery prywatno-intymnej, to nie ma tego samego poziomu jasności odnośnie do strumienia *metadanych*, które jednak zawierają równie szczegółowy i równie intymny obraz mojego życia.

Prawdziwie rewolucyjna jest wykluwająca się stopniowo w świecie Wielkiej Technologii idea, że dane same w sobie są tak naprawdę... zbędne. Są ciężkie, nieporęczne, zaszumione. Jest ich, co gorsza, *naprawdę* dużo. Choć lubimy myśleć o gigantach technologicznych jako o wszechpotężnych molochach, w rzeczywistości tempo produkowania dysków twardych, budowania centrów danych i kładzenia kabli podmorskich jest ograniczone – podobnie jak możliwość przetwarzania danych – tymczasem tempo generowania danych przez nas, użytkowników, jest wariackie. Spektakularną wczesną jaskółką tego procesu była historia huraganu Sandy, który pod koniec października 2012 roku wyrządził gigantyczne straty w Ameryce Północnej. Gdy dotarł do wybrzeża USA, na samym tylko Instagramie w ciągu niecałych dwóch dni opublikowano 1,3 miliona zdjęć tego huraganu, średnio 10 na sekundę.

Prawdziwego zalewu danych możemy się jednak spodziewać wraz z upowszechnianiem się „**internetu rzeczy**” (Internet of Things, IoT) – bransoletek medycznych, zegarków elektronicznych, lodówek, telewizorów i odkurzaczy typu „smart”, samochodów autonomicznych, domowych systemów monitoringu i tak dalej. Warto wrócić do wspomnianej już wyżej statystyki, że liczba takich urządzeń ma w 2023 roku wynieść 29,3 miliarda. Każde z nich produkuje tymczasem dane – głównie na temat nas i naszego otoczenia. Szacuje

się, przykładowo, że pojedynczy samochód autonomiczny będzie generował około 5–30 TB danych dziennie, głównie w postaci obrazu z licznych kamer wideo. Co z tymi wszystkimi danymi?

Oj, nie bójcie się, nie przepadną. Google Cloud już oferuje specjalną usługę adresowaną wyłącznie dla producentów samochodów autonomicznych, pozwalającą im gromadzić wszystkie dane generowane przez pojazdy „w chmurze” (czytaj: w centrach danych Google). Cóż jednak po niezinterpretowanych danych? Przypuśćmy, że któryś z partnerów Google, rządowych czy prywatnych, mający zezwolenie na dostęp do tych danych, zażyczy sobie informacji o wszystkich przypadkach, gdy którykolwiek z samochodów autonomicznych krążących po świecie albo „inteligentnych domofonów” zarejestrował twarz jakiejś określonej osoby. Analizowanie wszystkich danych za każdym razem, gdy pojawi się taka potrzeba, to nonsens – wszak wyszukiwarka Google też nie sprawdza dla nas „ręcznie” każdej strony w internecie, gdy tylko zadamy jej jakieś pytanie. Kluczowa jest więc „analiza na bieżąco”. Czyli: generowanie metadanych.

Doskonałej demonstracji tej zasadniczej idei dostarcza CERN, gdzie od lat 50. generowane są niewyobrażalne ilości danych. Jeden tylko zderzacz, LHC, generuje w ciągu sekundy 1 petabajt danych. Gdyby zachowywać wszystkie te dane, w ciągu 24 godzin zajęłyby łącznie 86 eksabajtów, czyli po dwóch tygodniach wypełniłyby wszystkie centra danych na świecie. W rzeczywistości dane te przechodzą jednak od razu przez szereg filtrów, a ostatecznie zapisuje się (w dużym stopniu na taśmach magnetycznych!) „tylko” 1 petabajt dziennie (29 czerwca 2017 roku centrum danych CERN przekroczyło granicę 200 PB pojemności). Nawet jednak i te petabajty pozostają bezużyteczne, jeśli nie podda się ich analizie. Każdy krok takiej analizy prowadzi ku coraz mniejszej ilości coraz bardziej „rozumiałej” i coraz wyżej zinterpretowanej treści: od „surowych danych” przez „dane zinterpretowane”, potem „informację”, aż ostatecznie „wiedzę” (czy nawet „mądrość”?). Przypomnijmy sobie rok 2012, kiedy to obwieszczono: „Bozon Higgsa istnieje”. Te trzy słowa to kondensat wielu petabajtów surowych danych.

Choć więc giganci technologiczni z przerażającą konsekwencją nie usuwają żadnych danych, które kiedykolwiek wpadły im w ręce, od samych surowych danych cenniejsza jest wygenerowana na ich podstawie informacja i wiedza. Przykładowo, na początku czerwca 2021 roku w nowej wersji polityki prywatności TikToka znalazła się [zgoda](#) na analizę wszystkich filmów tworzonych przez użytkowników

tej aplikacji i automatyczne generowanie na ich podstawie *faceprints* i *voiceprints*. Są to cyfrowe identyfikatory twarzy i głosu – coś w stylu ich ogólnego modelu pozwalającego na identyfikację. Użytkownicy zgadzają się też na identyfikację „przedmiotów widocznych na obrazie i krajobrazów w tle, twarzy i części ciała oraz ich cech, charakteru ścieżki dźwiękowej oraz tekstu wypowiedzianych w trakcie nagrania słów”^[80]. Instagram i Facebook chwala się, że dzięki automatycznej analizie umieszczanych na tych platformach obrazów, automatycznie powstają opisy zdjęć dla osób niedowidzących^[81]. To szczytny cel, jednak oznacza w praktyce potwierdzenie, że każde zdjęcie jest analizowane nie tylko pod kątem występujących na nim osób, ale również i obiektów. W 2018 roku również Google ogłosiło^[82], że jest w posiadaniu algorytmu zdolnego, dzięki mocy głębokiego uczenia maszynowego, do odtworzenia trójwymiarowej „mapy” przestrzeni na podstawie płaskich obrazów generowanych przez zwykłego smartfona.

A oto inny klarowny przykład na to, jaka jest różnica między danymi a tym, co można na ich podstawie wygenerować. W 2019 roku Facebook zaczął eksperymentować z identyfikacją użytkowników poprzez *video selfie* – krótki, kilkusekundowy film przedstawiający głowę z różnych stron^[83]. Oficjalnym pretekstem jest walka z fikcyjnymi kontami, a przedstawiciele firmy zapewniają, że filmy te będą usuwane natychmiast po zakończeniu procesu identyfikacji. Nie w samym pliku wideo tkwi jednak moc przesyłanych Facebookowi danych. Jeżeli na podstawie tego filmu stworzony zostanie trójwymiarowy model głowy – co dla dzisiejszych algorytmów jest już zadaniem niemalże rutynowym – to sam film można spokojnie usunąć. Model głowy, przypisany do konkretnego użytkownika, jest od niego znacznie cenniejszy.

Inaczej mówiąc, „surowe dane” przestają być potrzebne – i rzeczywiście mogą być one spokojnie usunięte, jeżeli zostanie to wymuszone – jeśli giganci technologiczni mogą swobodnie generować dowolnie szczegółowe opisy tych danych. Tego typu opisy – miejsc, osób, twarzy, przedmiotów – mogą być zaś przechowywane i analizowane do woli, zwykle pod dodatkowym pretekstem trenowania algorytmów i wymogów prawnych.

Oczywiście, możemy być pewni, że surowe dane również nieprędko znikną z serwerów. Przykładowo, „usunięcie konta” na Facebooku oznacza początkowo tylko „dezaktywację”, która w praktyce polega na dopisaniu słowa „nieaktywne” do gigantycznego repozytorium danych z twoim imieniem i nazwiskiem. Jeśli jednak uprzemy się, aby konto

naprawdę usunąć, Facebook – po 30-dniowym okresie, kiedy będzie uparcie namawiał nas na powrót na łono rodziny, kusząc ulubionymi treściami – faktycznie usuwa wszystko. Chyba. Mark Zuckerberg, zapytany w 2018 przez amerykańskiego senatora Deana Hellera, czy dane użytkowników naprawdę są usuwane po usunięciu konta, najpierw odparł, że nie jest pewien, a następnie powiedział tak: „Wiem, że staramy się je usuwać rozsądnie szybko. Mamy wiele złożonych systemów i trochę zajmuje przepracowanie tego wszystkiego”^[84]. Dziś Facebook twierdzi, że usunięcie danych użytkownika zajmuje 90 dni, przy czym „kopie niektórych treści mogą pozostać, jednak oddzielone od identyfikatorów osobistych”, a ponadto „informacje mogą być też przechowywane dla celów prawnych, związanych ze złamaniem warunków użytkowania, albo zapobiegania szkodom”^[85]. Krótko mówiąc, „może usuniemy coś”.

Nikt zaś tak naprawdę nie liczy na to, że usunięciu ulegną metadane albo – mówiąc jeszcze ogólniej – wszystko to, co powstało dzięki naszym danym. Jeżeli – czego nikt nie jest w stanie potwierdzić ani wykluczyć – istnieje gdzieś „centralne repozytorium” zawierające przypuszczalne położenie wszystkich użytkowników Facebooka od momentu założenia przez nich konta, a także szereg innych informacji na ich temat, to możemy być pewni, że dane te nie ulegną usunięciu. Nigdy.

Moment, a jakie właściwie „inne informacje na nasz temat” da się wyciągnąć, jeśli podejdziesz się do terabajtów naszych danych z podejściem „à la CERN”?



Automatyczny test osobowości

Wielka Piątka (*Big Five*) to jeden ze standardowych modeli osobowości w psychologii. Chociaż zajmujący się bliżej osobowością naukowcy wiedzą, że kryje ona w sobie znacznie więcej niż tylko 5 czynników, Wielka Piątka stosowana jest nie tylko w tysięcznych testach online, ale również w rzeczywistej praktyce badawczej i klinicznej jako dogodne „pierwsze przybliżenie”, z jakim właściwie z grubszą człowiekiem mamy do czynienia. Każdy z pięciu głównych czynników dość łatwo zrozumieć poprzez składające się na niego bardziej szczegółowe elementy, nazywane

technicznie „skalami”. W każdym przypadku jest to 6 skal (co daje ich łącznie 30), ja przytoczę tu tylko ich mały wybór, dla szybkiej orientacji:

- neurotyczność (niektóre skale: lęk, depresja, nadwrażliwość, nieśmiałość),
- ekstrawersja (towarzyskość, serdeczność, poszukiwanie doznań),
- otwartość na doświadczenie (wyobraźnia, estetyka, działanie),
- ugodowość (zaufanie, prostolinijność, skromność, ustepliwość, altruizm),
- sumiennność (skłonność do porządku, obowiązkowość, samodyscyplina, rozważa).

Początkowo do oceny osobowości danej osoby według modelu Wielkiej Piątki stosowano kwestionariusz NEO PI-R składający się z 240 zdań; później przedstawiono też wersje skrócone: 120-zdaniową, a nawet 60-zdaniową. Około 2017 roku postępy w głębokim uczeniu maszynowym oraz rosnąca dostępność danych z mediów społecznościowych sprawiły jednak, że jak grzyby po deszczu pojawiły się badania nad automatyczną „ekstrakcją” czyjegós modelu osobowości na podstawie jego aktywności w internecie^[64*].

Przewidywanie osobowości na podstawie czyichś słów okazało się względnie proste^[87]. (Oczywiście, słowo „proste” oznacza, że spędzono wiele miesięcy, pracowicie dopieszczając najbardziej zaawansowane spośród dostępnych wówczas technik uczenia maszynowego). Dane treningowe pochodziły od 250 osób, które wypełniły standardowy kwestionariusz psychologiczny, a następnie zgodziły się na anonimowe wykorzystanie ich statusów – czyli krótkich wpisów tekstowych – na Facebooku (średnio 40 na osobę) oraz kilku prostych liczb opisujących ich sieć społeczną (liczba znajomych, gęstość powiązań i tak dalej – są to parametry, które można w prosty sposób obliczyć, jeśli ma się dostęp do struktury czyjejs sieci społecznościowej). Na podstawie tych dwóch głównych źródeł danych udało się stworzyć model przewidujący osobowość danej osoby zgodnie z modelem Wielkiej Piątki z dokładnością 74% (gdzie 100% oznaczałoby całkowitą zgodność z kwestionariuszem tradycyjnym, a 0% wynik nieodróżnialny od losowego „strzelania”). To najlepszy wynik, jaki udało mi się znaleźć, choć nowe artykuły na ten temat ukazują się bezustannie, więc prawdopodobnie rekord ten został już pobity.

To jednak nie koniec. Trwają prace nad przewidywaniem osobowości wyłącznie na podstawie kliknięć – czyli, mówiąc językiem tego rozdziału: nie danych, a metadanych. To ambitniejsze zadanie, ponieważ algorytm nie ma dostępu do treści czyichś wypowiedzi, a tylko do wiedzy o tym, jakie inne treści zostały „zajakowane” lub „podane dalej”. Autorzy jednego z takich badań^[88] posłużyli się danymi od 92 tysięcy użytkowników Facebooka, którzy wypełnili test Wielkiej Piątki w aplikacji myPersonality, a ponadto wyrazili zgodę na anonimowe użycie danych o ich zachowaniu na Facebooku dla celów badawczych. Danymi treningowymi były więc pary test osobowości–lista „zajakowanych” treści. Autorzy posłużyli się ponadto sprytnym algorytmem, który opisał wszystkie ląjkowane treści przy pomocy automatycznie wygenerowanych 600 kategorii. Efektywnie więc, każdy badany użytkownik analizowany był nie tyle ze względu na listę polubionych treści, co raczej skłonność do „ląjkowania” każdej z tych 600 kategorii. Cóż to za kategorie? Gdyby listę tę układał człowiek, prawdopodobnie nosiłyby one nazwy typu „kulinaria”, „treści prolewicowe”, „memy z kotami” czy „historie osobiste o zdradzie”. Listę tę jednak „układał” algorytm, a jego jedynym celem było stworzenie takich 600 „wymiarów”, które będą najskuteczniej różnicować ludzi ze względu na ich osobowość. Jest to więc procedura bardzo pokrewna „wektoryzacji słów” opisaną w ramce na stronie 99: całe uniwersum treści zostało sprowadzone do kilkuset abstrakcyjnych parametrów. W tym przypadku „punktami” owej 600-wymiarowej przestrzeni są jednak ludzie.

Tak czy inaczej, projekt zakończył się umiarkowanym sukcesem. Dokładność oszacowań pięciu składników Wielkiej Piątki wahała się w przedziale 22–38%. Jak

wyżej, trzeba jednak pamiętać, że niemal na pewno nie jest to najlepszy wynik możliwy do uzyskania taką metodą. Tempo rozwoju tej dziedziny jest spore, a poza tym można rozsądnie przypuszczać, że te naprawdę najlepsze algorytmy nie są opisywane w publicznie dostępnej literaturze naukowej, lecz są przedmiotem patentów lub należą do, mówiąc technicznie, „tajemnicy przedsiębiorstwa” – programiści Google’a i Facebooka po prostu nie mają najmniejszej ochoty zdradzać ich światu.



Facebook cię zdiagnozuje

W maju 2020 roku ukazał się fenomenalny artykuł przeglądowy^[89] na temat pozyskiwania wiedzy o „sygnałach biomedycznych i zachowaniach związanych ze zdrowiem” na podstawie treści osobistych publikowanych w mediach społecznościowych. Warto przyjrzeć się choćby samym tytułom artykułów, które tam cytowano. Po ich przejrzeniu czasami okazuje się, że metodologia lub skuteczność pozostawiają wiele do życzenia, ale samo istnienie tych badań jest już dostatecznie interesujące. To wgląd w to, jakiego typu dane na nasz temat można uzyskać na podstawie naszej aktywności w internecie. Oto skromna próbka:

- *Lingwistyczna analiza schizofrenii w postach na Reddit*^[90]. Autorzy posłużyli się uczeniem maszynowym, aby stworzyć „automatyczny detektor użytkowników ze schizofrenią”. Algorytm analizował wypowiedzi 159 użytkowników – średnio po 418 postów na osobę – którzy uczestniczyli w grupach tematycznych („subredditach”) typu „r/schizophrenia” i twierdzili w postach, że zdiagnozowano u nich schizofrenię – oczywisty słaby punkt badania, z którym naukowcy jednak się nie kryją. Dla porównania zebrano podobną objętość wypowiedzi 159 osób z losowo wybranych grup tematycznych. Tekst pochodzący od każdego użytkownika przeanalizowano przy pomocy narzędzia o nazwie LIWC, które wyznacza dla danego zbioru tekstów proste „techniczne” parametry (typu: częstość stosowania różnych spójników, typ używanych znaków interpunkcyjnych, częstość używania pierwszej, drugiej i trzeciej osoby, czasu przeszłego, teraźniejszego oraz przyszłego i tak dalej), ale też dzieli użyte w tekście słowa na rozmaite kategorie tematyczne (słowa związane z emocjami pozytywnymi i negatywnymi, dotyczące religii, myślenia, relacji społecznych). Tego typu „profile językowe” był już danymi treningowymi dla algorytmu uczenia maszynowego, który następnie przetestowano na nowej próbie użytkowników. Autorzy chwalą się dokładnością 81,6%.
- *Monitorowanie potencjalnych oddziaływań leków i narkotyków za pośrednictwem analizy sieciowej użytkowników Instagrama*^[91]. Metodologia podobna, jednak mocno oparta na tagach wpisywanych przez użytkowników Instagrama, którzy uwielbiają – jak się okazuje – publicznie chwalić się przyjmowanymi przez siebie lekami. Ba, analizowane były również zdjęcia, na których fiołki z lekami prezentowane są często dumnie na pierwszym planie. Ludzie nie kryją się też

przesadnie z paleniem marihuany. Ostatecznie udało się uzyskać sporą bazę danych osób, które prawdopodobnie przyjmowały popularne leki na depresję oraz podgrupę tych, którzy równolegle stosowali jakiś inny lek lub palili marihuanę. Następnie przeanalizowano opublikowane przez nich treści, łącznie pochodzące z ponad 5 milionów postów, w poszukiwaniu słów mogących świadczyć o niepożądanych skutkach ubocznych. Potwierdzono wiele znanych producentom i lekarzom skutków ubocznych i niebezpiecznych oddziaływań między lekami, a także zidentyfikowano kilka nowych przypuszczalnych oddziaływań, jak na przykład częstsze od przypadkowego występowanie słów związanych ze skokami ciśnienia krwi (zwłaszcza „nadciśnienie”) u osób, które równocześnie piszą o łuszczycy i przyjmują metforminę, stosowaną w leczeniu cukrzycy typu II.

- *Ludzkie cykle płciowe są sterowane kulturowo i są zgodne z nastrojami społecznymi*^[92]. Tym razem zbadano skłonność ludzi do wyszukiwania poprzez Google słów związanych z seksualnością oraz używanie odpowiednich słów kluczy na Twitterze. Jednym z celów tego typu badań jest wykrzycie, na jakim etapie cyklu miesięczkowego są w danym momencie kobiety korzystające z tych usług. Przy okazji okazało się, że zainteresowanie ludzi seksem rośnie w czasie świąt, zwłaszcza świąt Bożego Narodzenia u chrześcijan i Id al-Fitr, czyli Święta Przerwania Postu, kończącego Ramadan, u muzułmanów. Udało się też zaobserwować wzrost liczby narodzin 9 miesięcy po tego typu okresach, co potwierdza, że owo ożywienie nie kończy się na klepaniu w klawiaturę.

Co z tego wszystkiego wynika? I co przyniesie przyszłość?

Trudno powiedzieć.

Wiemy mniej więcej, co nas otacza: niepowstrzymany strumień danych, spływających najpierw lokalnymi strumyczkami, a potem wielkimi rzekami do zbiorników, których właściciele są bardzo niechętni idei usuwania z nich czegokolwiek, a każdy trafiający do nich bit łączywie analizują, generując na jego podstawie szczegółowe profile osobiste, zawierające miliony „punktów danych”. To wiemy – tylko co dalej?

Niektórzy zapaleńcy próbują wyrwać się z Systemu. Tak zwani **degooglerzy** starannie rekonfigurują swoje laptopy i smartfony, aby nie znalazł się w nich choćby i większy segment kodu stworzony w Google, a następnie programowo odmawiają korzystania z usług tej firmy. Wymaga to pewnego otrząskania technologicznego i kilku poświęceń, zwłaszcza jeśli przywykło się do Gmaila, YouTube czy Google Maps. Tworzone przez degooglerów poradniki stają się jednak z miesiąca na miesiąc coraz bardziej przyjazne dla użytkowników i rośnie zainteresowanie promowanym przez nich ruchem. Powstają też bliźniacze inicjatywy typu „deApple”, „deAmazon” czy „deFacebook”.

W świetle faktów, które omawiam w ramce na stronie 156, można powątpiewać, czy program ten dałoby się przeprowadzać naprawdę konsekwentnie – w bazy danych Google trafiamy również za sprawą „drugiego stopnia pokrewieństwa”. Całą procedurę „odgooglowania” ostatecznie może więc trafić szlag, kiedy nasz pracodawca wymusi na nas założenie konta na Google Hangouts albo będzie koordynował naszą pracę poprzez Kalendarz Google. Co więcej, nasi bliżsi i dalsi znajomi mogą wciąż z uporem maniaka oznaczać nas na zdjęciach na Facebooku. Choćbyśmy nawet całkowicie zrezygnowali z urządzeń elektronicznych – popuścimy może nieco wodze fantazji – sfotografować nas może któryś z pojazdów uaktualniających zdjęcia uliczne dla Google Street View, a sfilmować któryś z pojazdów autonomicznych, nie mówiąc już o sieci monitoringu miejskiego. Chcecie się jeszcze trochę postraszyć? W kwietniu 2020 roku wyszło na jaw^[93], że w internecie dostępnych jest 8,6 miliona zapisów z automatycznego systemu rozpoznawania tablic rejestracyjnych, stosowanego w brytyjskim mieście Sheffield. Na podstawie zgromadzonych danych można było dowiedzieć się, w zasięgu której spośród około setki kamer monitoringu miejskiego znajdował się określony pojazd w zadanym czasie. Ponieważ nie ma żadnych wiarygodnych informacji na temat realnego stopnia wymiany informacji pomiędzy firmami typu Google, Amazon czy Facebook oraz instytucjami państwowymi, tylko od indywidualnego poziomu podejrzliwości zależy, jak gęste powiązania między nimi sobie wyobrazimy.

Zmaganie to tym ostrzej ujawnia jednak, w jak ciekawych czasach żyjemy (obyśmy nie żyli w *zbyt* ciekawych czasach). Trzeba by potężnego wysiłku i sporych wyrzeczeń, aby nie przechowywano i nie analizowano gigantycznych ilości danych na nasz temat.

Prawdopodobnie protest przeciwko temu stanowi rzeczy byłby znacznie silniejszy, gdyby ktoś w końcu wyartykułował naprawdę przekonującą i naprawdę przerażającą wizję tego, co można uczynić z kompletem danych na czyjs temat. Jest wiele „obietujących” tropów – kwestie intymności, wstydu i narażenia na śmieszność; wrażliwych danych medycznych; ujawniania mniejszych i większych przestępstw, szantażu... wciąż nie jest to jednak coś, co pchnęłoby masy ludzkie na ulice i kazało im publicznie palić swoje telefony. A może przesadzamy i z jakiegoś powodu okaże się to lękiem nieusprawiedliwionym?

Twoje ciało

Marzenia o wiecznym zdrowiu i życiu kojarzą się raczej mitologicznie niż naukowo – ale może tylko dlatego, że nie doceniamy medycyny, biotechnologii, inżynierii genetycznej? Tymczasem tylko od XIX wieku przewidywana długość życia statystycznego przedstawiciela *Homo sapiens* wzrosła z nieco ponad 30 do prawie 70 lat^[94]. Inne źródło podaje, że średnie tempo przyrostu długości życia to 2,5 roku na dekadę^[95]. Pytanie za sto punktów brzmi, czy trend ten można utrzymywać dowolnie długo? Zanim przyjrzymy się najnowszym postępom w wydłużaniu życia, poświęćmy odrobinę uwagi statystykom dotyczącym, no cóż, śmierci.

W środowisku badaczy zajmujących się długowiecznością panuje pełna różnorodność opinii w temacie możliwości nieograniczonego przedłużania życia. Jedni uważają, że istnieje „sufit”: jedyne, na co możemy liczyć, to zbliżanie się z długością życia do pewnej granicy, głównie sprawiając, że coraz mniej osób będzie umierać „za wcześnie”. Bądź co bądź, tak właśnie osiągnęliśmy obecny wspaniały wynik: w przeważającym stopniu za wspomniany wzrost średniej odpowiada spadek współczynnika umieralności niemowląt i dzieci.

Obecnie na 1000 urodzonych na świecie dzieci 27 umiera w ciągu pierwszego roku życia („śmiertelność niemowląt”), a 45 przed 15 rokiem życia („śmiertelność dzieci”)^[65*]. Inaczej mówiąc, 95,5% ludzi dożywa swoich 15 urodzin. Z czysto ludzkiego punktu widzenia każda z tych śmierci to oczywiście tragedia, jednak statystycy i lekarze patrzą na te liczby z radosnym niedowierzaniem. Wszystkie badania sięgające w przeszłość naszego gatunku^[96] prowadzą do wniosku, że „za dawnych, dobrych lat” śmiertelność niemowląt wynosiła około 250–300/1000 (czyli 25–30%), a śmiertelność dzieci około 450–500/1000 (czyli 45–50%). Co ciekawe, do tych samych liczb prowadzą badania dzisiaj żyjących społeczności łowiecko-zbierackich, analizy historyczne oraz głębsze, paleontologiczne na temat neandertalczyków i wczesnych *Homo sapiens*, a także porównawcze wyniki z populacji szympanсів. Wygląda na to, że od zarania dziejów

tylko mniej więcej połowa osób dożywała wieku dorosłego, podczas gdy dziś jest to niemal 96%!

Statystyki te, jakże krzepiące, bywają jednak przytaczane na dowód tego, że za przyrost średniej długości życia nie odpowiada podnoszenie się „sufitu”, tylko eliminacja śmierci „przedwczesnej”. Pytanie jednak, gdzie jest ów sufit? Niemal wszyscy zainteresowani podają wartość 120 lat. Najdłuższe wiarygodnie zarejestrowane życie należało do Francuzki Jeanne Calment, która zmarła 4 sierpnia 1997 roku w wieku 122 lat i 164 dni (żyła łącznie 44 724 dni). Był to jednak ewenement: druga na liście rekordzistów Sarah Knauss przeżyła 119 lat i 97 dni – w całej historii wiarygodnych zapisów narodzin był więc tylko jeden przypadek przekroczenia owej „magicznej” granicy. Czyżby więc przyszłość miała wyglądać tak, że ludzie będą po prostu coraz częściej i w coraz lepszym zdrowiu żyć 100 i 110 lat, jednak nikt nigdy nie będzie żył, przypuścmy, 130 lat? To by nie był straszny świat, jednak my mamy oczywiście apetyt na więcej.

W lutym 2021 roku ukazał się ciekawy artykuł^[97] na temat przyszłych trendów w długowieczności. Jego autorzy zwracają w pierwszej kolejności uwagę na fakt, że tzw. „przewidywana długość życia” w momencie urodzenia z zasady nigdy się nie sprawdza – ze względu na rozwój medycyny. Przykładowo, w roku mojego urodzenia (1985) przewidywana długość życia polskich mężczyzn wynosiła 66 lat. W ciągu ostatnich 36 lat medycyna poszła jednak do przodu, tak więc (nawet po wzięciu dodatkowej poprawki na to, że należę do tych szczęśliwców, którzy dożyli swoich 36. urodzin) moi równolatkowie mogą liczyć na znacznie więcej niż 30 lat dalszego życia (GUS podaje, że jest to nieco ponad 40 lat).

Biorąc to pod uwagę, autorzy tekstu szacują, że większość dzieci urodzonych w krajach wysoko rozwiniętych w ciągu ostatnich 20 lat dożyje swoich setnych urodzin – mimo że średnia przewidywalna długość życia w tych krajach, wyczytana po prostu z tabel statystycznych, nie wynosi 100 lat. Po drugie, zauważają też, że wiek rekordzistek i rekordzistów w „tabeli ligowej” długowieczności powoli, ale stale się podnosi, w uśrednionym tempie 1,5 miesiąca na rok, czyli nieco ponad rok na dekadę. Czyżby więc „sufit” mimo wszystko się podnosił?

Oczywiście, wszystkie te obliczenia mogą okazać się bezużyteczne, jeśli nastąpi jakiś gigantyczny przełom w medycynie lub – realistyczniej – szereg ząbwiąjących się „przełomików”. Z drugiej

strony, statystyki mogą też trafić do kosza choćby za sprawą nowych patogenów jak bakterie lekooporne albo wirusy.

To typowy problem z przewidywaniem przyszłości, o którym musiałbym w tej książce przypominać chyba na każdej stronie – pozwólcie mi na ograniczenie się tylko do jednej demonstracji, w czym rzecz. Na początku XX wieku istnieli prawdopodobnie futurologi, próbujący przewidzieć, jak szybko będzie można podróżować za 100 lat. Futurologi tacy mogliby się odwoływać do istniejących rekordów: od parowozu Richarda Trevithicka, który w 1804 osiągnął 8 km/h, przez Rakietę Roberta Stephensona, która w 1829 roku rozpedziła się do 46 km/h, aż po świeżutkie wówczas doniesienia o eksperymentalnym elektrowozie firmy Siemens & Halske, który latem 1901 roku pomknął z zawrotną prędkością 160 km/h. Futurologi owi mogliby „nanizać” te liczby na linię, konkludując, że człowiek w XXI wieku będzie mógł przemieszczać się nawet z prędkością 320 km/h. Gdyby jednak wstrzymali się z publikacją zaledwie 3 lata, być może zastanowiłby ich niepozorny Flyer braci Wright, którego duchowi potomkowie już w latach 20. przekroczyli prędkość 500 km/h.

Nie bójmy się marzyć! Może naprawdę technologia przyszłości pozwoli nam żyć dwieście lat?

Zerknijmy ku kilku tematom we współczesnej „inżynierii medycznej”, które mogą w XXI wieku porwać nas ku niebiosom. Podzieliłem je na trzy kategorie tematyczne:

- naprawianie psującego się ciała w nieskończoność (rozdział IV.1);
- naprawianie i „projektowanie” ludzi metodami genetycznymi (rozdział IV.2);
- „ulepszanie” człowieka i rozszerzanie jego możliwości (rozdział IV.3).

IV.1. Naprawiamy w nieskończoność

Użyłem wyżej określenia „inżynieria medyczna”, które jest bardzo nieostre, trochę w stylu „sztucznej inteligencji”. Tak naprawdę, kiedy pojawia się temat „naprawiania ciała”, w jednej chwili otwierają się setki szufladek: stomatologia i kosmetologia, onkologia, kardiologia i ortopedia, nanotechnologia, terapia genetyczna, drukowanie 3D biomateriałami, hodowle komórkowe, diagnostyka z użyciem sztucznej inteligencji, medycyna spersonalizowana... i tak dalej, i dalej, i dalej...

Gdybym miał być zupełnie szczery, to musiałbym powiedzieć tak: plan „dowolnie długiego” przedłużania zdrowia i życia ludzkiego oznacza w praktyce robienie *wszystkiego* tego, co robi się dzisiaj w medycynie, coraz lepiej. Gdzieś na świecie są teraz ludzie pracujący nad leczeniem kamicy nerkowej, którzy sprawiają, że za 20 lat będzie ono skuteczniejsze i mniej bolesne, co przybliży ku realizacji zarysowany w tym rozdziale plan o kolejny milimetr.

Nikt poważny nie wierzy w to, że wynaleziony zostanie „złoty pocisk”, który za jednym zamachem rozwiąże wszystkie problemy medyczne (podobnie jak nie ma „po prostu sztucznej inteligencji”, która rozwiąże problem „myślenia maszynowego”). Mało tego, nawet sformułowania na pozór szczegółowe, jak „lek na raka”, budzą tylko politowanie ekspertów, którzy wiedzą, że jest to zbyt ogólne sformułowanie problemu. Jaki jest więc plan gry na ten rozdział? Plan kompletnego przeglądu stanu i perspektyw medycyny oczywiście odpada. Zamiast tego, postanowiłem pomówić o kilku wybranych, szczególnie spektakularnych osiągnięciach z ostatnich lat i dekad, które mają potencjał, aby poważnie wpłynąć na wiele różnych dziedzin medycyny.

Zacznijmy od idei względnie starej, ale wyraźnie przyspieszającej w ostatnich latach – o odtwarzaniu uszkodzonych, starych lub zniszczonych organów i tkanek *in vitro*.

IV.1.1. Inżynieria tkankowa



Komórki macierzyste

Istotę komórek macierzystych (*stem cells*) najłatwiej zrozumieć po przyjrzeniu się rozwojowi organizmu człowieka. Kiedy już pszczołka zapyli kwiatek, powstaje jedna komórka, zwana zygotą – pierwsza komórka nowego organizmu. Po niedługim czasie

zaczyna się ona dzielić, najpierw tworząc mikroskopijną „malinkę” nazywaną blastulą, która potem – w takt dzielących się bezustannie komórek – przechodzi przez wariackie embriologiczne origami: zwija się i skręca, rozciąga i nadmuchuje, tu się robią wypustki, tam wpuklenia... aż ostatecznie mamy do czynienia z kilkumiesięcznym płodem, który z grubsza przypomina już dorosłego człowieka. Historia jest względnie znana, spójrzmy jednak na nią z perspektywy znawcy i konesera komórek.

Dojrzały organizm składa się z kilkuset różnych typów komórek, od płaskich komórek naskórka, przez drzewkowate neurony, aż po okrągłutkie adipocyty, czyli komórki tłuszczowe; łącznie kilkadziesiąt bilionów sztuk. Wszystkie one są potomkami zygoty – nowe komórki nie powstają wszak inaczej niż poprzez podział jakiejś innej komórki. Jak to możliwe? Pierwsze pokolenia komórek, na etapie „malinki” i nieco później, pozostają całkowicie „niedookreślone” i czekają tylko na odpowiedni sygnał – są to głównie sygnały chemiczne, ale również mechaniczne – aby zacząć zamieniać się w komórki skóry, mózgu, nerki czy płuc. Tego typu niezróżnicowane komórki to **embrionalne komórki macierzyste**. Owa przemiana nie zachodzi jednak z pokolenia na pokolenie, a raczej następuje poprzez stopniowe „doprecyzowanie” losu danej linii komórek.

Przypuśćmy, że śledzimy jedną, wybraną „komórkę matkę”, która jest jeszcze całkowicie niedookreślona i ma w sobie potencjał, aby stać się dowolną komórką dorosłego organizmu (jest więc **pluripotencjalna**). Jedno z jej dzieci lub wnucząt pod wpływem odpowiednich substancji chemicznych (o sympatycznych nazwach typu Gooseoid czy Foxa-2) może zostać „odrobinę doprecyzowane” i stać się na przykład *komórką zarodkową ektodermy*, czyli warstwy embrionu, z której powstanie później skóra i układ nerwowy. Taka komórka będzie następnie dalej dzielić się, a ta „gałąź” jej potomstwa, przykładowo, która zostanie nasączona produkowanymi w przedniej części embrionu związkami (o jeszcze sympatyczniejszych nazwach jak Chordin czy Noggin) stanie się ostatecznie *komórkami zarodkowymi tkanki nerwowej (neural stem cells)*. Te z kolei komórki są już tylko o włos od swojego ostatecznego przeznaczenia: wystarczy jeszcze tylko trochę szturchnąć je chemicznie, aby przemieniły się na przykład w neuron, astrocyt albo którykolwiek z innych typów dojrzałych komórek budujących układ nerwowy. To **multipotencjalne komórki macierzyste** – mające kilka, może kilkanaście dalszych możliwości rozwojowych. Ostatecznie i te ulegną „doprecyzowaniu”, a dorosła, dojrzała, ustalona komórka jest już „po prostu sobą”.

Dobłą metaforą dla komórek macierzystych jest rozwój zawodowy pojedynczego człowieka w „organizmie społecznym” – a przynajmniej w bardzo karnym, ale i równościowym społeczeństwie. Małe dziecko, które jeszcze niczego nie umie, to taki „pluripotencjalny przedszkolak”. Z czasem wysyłamy go do „szkoły majsterkowiczów” albo „szkoły gryziopiórków”, albo „szkoły sportowo-siłowej” – wykształcony w niej „multipotencjalny nastolatek” ma przed sobą nieco zawężoną ścieżkę kariery. Ostatecznie wybrany mu zostaje określony zawód, przy którym pozostaje już do śmierci.

Uwaga, pierwszy haczyk. „Macierzystość” nie kończy się w łonie matki. Nawet w dorosłym organizmie ludzkim pozostają rozmaite populacje „uśpionych” komórek macierzystych, zdolnych do produkowania określonego rodzaju potomstwa – trochę tak, jak gdyby społeczeństwo planowo utrzymywało grupę bezrobotnych multipotencjalnych „majsterkowiczów”, aby we właściwym czasie po szybkim przeszkoleniu pchnąć ich na front robót, gdyby nagle zabrakło ślusarzy albo stolarzy. To **komórki macierzyste dorosłego organizmu (adult stem cells)**. Najbardziej znane są *komórki macierzyste szpiku (hematopoietic stem cells)*, występujące w szpiku kostnym większości kości. Są multipotencjalne – ich potomstwo może stać się ostatecznie każdą komórką krwi: od czerwonych krwinek, przez płytki, aż po leukocyty. Ich identyfikacja w latach 70. XX wieku spotkała się początkowo ze sporym

niedowierzaniem, podobnie jak kolejne podobne odkrycia – utrudnione przez fakt, że organizm pilnie strzeże swoich komórek macierzystych, ukrywając ten bezcenny skarb głęboko w tkankach.

Komórki macierzyste serca (cardiac stem cells) występują w drobnych skupiskach rozsianych w mięśniu sercowym, głównie w ścianach przedsionków, gdzie biorą udział w naprawie tkanek po uszkodzeniu. Nieco łatwiej dostępne są *komórki macierzyste mieszka włosowego*, mogące po „wybudzeniu” generować rozliczne komórki skóry. Do dziś nie jest jasne, ile tego typu populacji występuje w organizmie ludzkim, choć coraz częściej podejrzewa się, że praktycznie każda tkanka ma własną populację tzw. **komórek progenitorowych**, czyli komórek macierzystych „prawie dojrzałych”, będących dosłownie o krok lub dwa od stanu dojrzałego. Szczególnym źródłem komórek macierzystych jest też krew pepowinowa, w której znajdują się znaczne ilości tzw. *mezenchymalnych komórek macierzystych* – to bardzo popularne źródło w badaniach eksperymentalnych.

Haczyk drugi. Istnieje możliwość przekwalifikowania się. W zwykłych warunkach panujących w organizmie neuron nie zamieni się w komórkę wątroby, jednak od czego odrobina pomysłowości – i oczywiście doskonała znajomość chemicznych sygnałów wywołujących różnicowanie. W 2006 roku Shinya Yamanaka wykazał, że fibroblasty – poutykane w tkance łącznej komórki, produkujące m.in. kolagen – po potraktowaniu ich zaledwie 4-składnikowym koktajlem chemicznym zamieniają się w komórki macierzyste. W ten sposób na arenie medycyny pojawiły się **indukowane pluripotencjalne komórki macierzyste (iPSC)**, za których wytworzenie Yamanaka otrzymał w 2012 roku Nagrodę Nobla. Dziś komórki iPSC są już standardowym narzędziem biotechnologii. Ponieważ fibroblasty ukryte są pod skórą, szybko przystąpiono do prac nad dogodniejszym źródłem. Już w 2008 roku opisano wywołanie („indukowanie”) macierzystości w keranocytach, pobranych z wyrwanego włosa; w 2010 roku z komórek krwi pobranej z palca; a w 2012 z komórek nabłonkowych obecnych w moczu. Oznacza to, że dziś można by utworzyć kolonię (indukowanych) komórek macierzystych konkretnej osoby bez kłopotliwego pobierania tkanek; ba, nawet bez wiedzy takiej osoby.

Czasem dla skutecznego indukowania macierzystości konieczne są modyfikacje genetyczne komórek „dorosłych”, a sama procedura „odróżnicowywania” jest niestety dość skomplikowana i nie zawsze udaje się uzyskać stopień „plastyczności” porównywalny choćby z komórkami embrionalnymi.

Dzisiaj wszystkie trzy główne źródła – embrionalne, dorosłego organizmu i indukowane – są stosowane w medycynie, każde obciążone swoimi wadami. Szczególnie kuszącym źródłem są embrionalne komórki macierzyste, dla pozyskania których nie trzeba włamywać się do szpiku czy mięśnia sercowego, albo żmudnie „odprogramowywać” zwykłych komórek – co nawet dziś, 15 lat po odkryciu Yamanaki, wcale nie jest prostym zadaniem. Komórki embrionalne są jednak, jak łatwo się domyślić, silnie obciążone etycznie – są to, mimo wszystko, dokładnie takie same komórki, z jakich rozwija się prawdziwy ludzki embrion.

Choć potencjał komórek macierzystych jest gigantyczny, postęp w medycynie następuje powoli. Przykładowo, naiwnie mogłoby się wydawać, że stanowią one oczywistą odpowiedź choćby na uszkodzenie rdzenia kręgowego. Przerwany rdzeń? Nic prostszego – wstrzykujemy w miejsce uszkodzenia trochę komórek macierzystych wraz z koktajlem „unerwiający” i... voilà! Nie brakuje oczywiście chętnych do przeprowadzania tego typu eksperymentów: niedawny przegląd^[98] wymienia 27 badań klinicznych nad leczeniem urazów rdzenia komórkami macierzystymi. Najstarsze z nich sięgają 2008 roku, a warto podkreślić, że badanie kliniczne, nawet pierwszej fazy^[66*], to już zaawansowany etap eksperymentów. Praktyka jest jednak, jak zwykle, bardziej skomplikowana i mniej świetlista od czystej teorii. Komórki macierzyste wydają się owszem wspomagać leczenie rdzenia kręgowego, ale nie we wszystkich przypadkach, nie bez żmudnej fizjoterapii i nie stuprocentowo. Dość szybko stało się też jasne, że „wstrzyknięcie” porcji komórek to trochę za mało: każda tkanka ludzka jest wysoce zorganizowana przestrzennie. Jeżeli chcemy zastąpić jakiś fragment organizmu ludzkiego nowymi komórkami, trzeba je najpierw odpowiednio uporządkować. W ten sposób badania nad komórkami macierzystymi sprzymierzyły się z inżynierią tkankową.



Inżynieria tkankowa

Termin „**inżynieria tkankowa**” obejmuje wszelkiego rodzaju starania, aby odtworzyć ludzkie tkanki, uszkodzone fizycznie lub w wyniku choroby. Korzenie inżynierii tkankowej tkwią w chirurgii, zwłaszcza plastycznej – czasem mówi się, że pierwszymi inżynierami tkankowymi byli chirurdzy wojskowi, pracownicy „łatający” rannych żołnierzy, czasem rekonstruując przy tym zniszczone organy. W połowie XX wieku ruszył program wszycia pacjentom materiałów syntetycznych, na przykład nylonu, aby pomóc ciału w regenerowaniu tkanek. Przypadki „proste”, kiedy to celem jest na przykład odtworzenie uszkodzonej kości, na dobre weszły już do medycyny^[99] –

brakujący fragment kości modeluje się najpierw w odpowiednim programie komputerowym, a następnie drukuje (najczęściej z materiałów ceramicznych, mineralnych i metalicznych) i wszczepia. To już dziś niemal rutynowa procedura, leżąca gdzieś na granicy między inżynierią tkankową, protetyką i transplantologią.

Sytuacja jest dużo bardziej skomplikowana, gdy celem jest odtworzenie złożonego, miękkiego organu, jak skóra, wątroba, płuco, nerka czy serce. W ogólności jest to problem jeszcze nierozwiązany, który można rozbić na dwa główne zadania:

- stworzenie odpowiedniego rusztowania (*scaffold*) – będzie to „martwa” część tkanki, organizująca ją przestrzennie i wspomagająca rozwój komórek;
- „zasiedlenie” (*seeding*) rusztowania odpowiednimi komórkami; eksperymentuje się też z bezpośrednim „wdrukowywaniem” komórek na etapie tworzenia rusztowania, jeśli stosowane są drukarki 3D.

Rusztowanie można stworzyć na różne sposoby. Jeżeli ma być ono względnie proste, na przykład mieć postać cienkiego arkusza (do przeszczepu skóry) albo cylindra (na przykład przy hodowaniu zamienników naczyń krwionośnych, jelita albo oskrzeli), czasem wystarczą nożyczki, igła i nitka – bo taka technologia w zupełności wystarczy, aby uszyć nylonową rurkę. Nieco bardziej zaawansowaną metodą jest **druk 3D** (zob. ramka na s. 81) przy pomocy jakiegoś materiału, który nie zostanie wykryty przez organizm jako ciało obce. Inną techniką jest posłużenie się prawdziwą, „żywą” tkanką, pochodzącą na przykład z organów, których nie da się wykorzystać do zwykłego przeszczepu. Z tkanki takiej wymywa się na różne sposoby komórki – to tzw. **decelularyzacja** – dzięki czemu pozostaje tylko międzykomórkowe rusztowanie: złożona struktura zbudowana z różnego rodzaju włókienek i cząsteczek, pełniąca w organizmie niedocenianą, ale szalenie istotną rolę. W ostatnich latach coraz częściej okazuje się, że ta niepozorna macierz międzykomórkowa w niemal „magiczny” sposób pomaga komórkom w prawidłowym rozwoju.

Jeżeli chodzi o komórki zasiedlające rusztowanie, to albo stosuje się komórki macierzyste, albo „dojrzałe” komórki określonego rodzaju lub rodzajów. Nawet najprostszy organ składa się z kilku, a typowy nawet z kilkudziesięciu odmian komórek, uporządkowanych przestrzennie z niezwykłą finezją. Co więcej, im lepiej poznajemy tajniki wzrostu organów, tym bardziej staje się jasne, że nie da się wyhodować rozmaitych komórek, przykładowo, wątroby, w szklanym naczyniu, a potem „wlać” je w odpowiednie rusztowanie, licząc na to, że powstanie z tego wątroba. Równie dobrze można by liczyć na to, że na setkach osobnych wysp wychowa się ślusarzy, młynarzy i urzędników, wpuści ich do opustoszałego miasta, a następnego dnia będzie w nim tętniło normalne życie społeczne, a fabryki zaczną produkować towary na eksport. W ostatnich latach popularne staje się hodowanie komórek macierzystych już w docelowym rusztowaniu, dzięki czemu, między innymi, od razu przybierają właściwy kształt, wymuszony przez środowisko, w jakim się znalazły.

Przyjrzyjmy się może kilku wybranym przykładom inżynierii tkankowej, żeby zorientować się z grubsza, na czym stoimy w 2021 roku.



Oskrzele z hodowli komórkowej^[100]

Zacznijmy od dość odległej historii, jak w świecie medycyny należałoby określić rok 2008. To wtedy jednak po raz pierwszy wszczepiono człowiekowi organ wyhodowany *in vitro* z komórek pacjenta. Oznacza to więc, że mówimy nie tylko o udanym eksperymencie, ale i przypadku monitorowanym przez wiele lat – czego nie można powiedzieć o świeżynkach z „tegorocznej” literatury naukowej.

Bohaterką cytowanego artykułu jest 30-letnia wówczas Claudia Castillo, Kolumbijka, która po ciężkim przypadku gruźlicy zgłosiła się do lekarza z poważnymi trudnościami z oddychaniem. Okazało się, że jej lewe oskrzele jest zapadnięte i praktycznie w ogóle nie przepuszcza powietrza do płuca. Postanowiono przeprowadzić eksperyment medyczny: pacjentkę sprowadzono do Barcelony, gdzie jej przypadkiem zajął się sztab ekspertów z rozmaitych europejskich szpitali i uniwersytetów.

Rozpoczęto od pobrania 7-centymetrowego odcinka tchawicy od 51-letniej kobiety, która niedługo wcześniej zmarła wskutek wylewu krwi do mózgu. Następnie przez 6 tygodni poddano go łącznie 25 cyklom decelularyzacji, aby usunąć z niego choćby najdrobniejsze ślady komórek dawczyni. Każdy taki cykl polegał na moczeniu w specjalnie przygotowanym roztworze rozpuszczającym tłuszcze i niszczącym DNA, a następnie dokładnym przemyciu wodą destylowaną. Kiedy już upewniono się, że rurka jest „czysta” i nie zostanie potraktowana przez organizm biorczyny jako ciało obce, przystąpiono do zasiedlania jej komórkami.

Ponieważ oskrzele jest strukturą warstwową i posiada zarówno nabłonek, jaki i komórki tkanki chrzęstnej (chondrocyty), od Kolumbijki trzeba było uzyskać obydwie typy próbek. W celu wyhodowania nabłonka podjęto dwie próby: jedna na podstawie materiału z jej własnego (prawego) oskrzela, a druga na podstawie próbki błony śluzowej z nosa. Komórki z tej pierwszej hodowały się lepiej i to one ostatecznie rozprawdzono po wewnętrznej stronie rusztowania. Chondrocyty uzyskano z mezenchymalnych komórek macierzystych ze szpiku kostnego, które nakierowano na „ścieżkę chondrocytową” standardową metodą znaną już wówczas z literatury. Te trafiły na zewnętrzną ściankę rusztowania. Sam proces „wysiewania” komórek trwał 96 godzin, a tchawica, zanurzona do połowy w odżywczej „zupce” o temperaturze 37°C, była dodatkowo stale powoli obracana wzdłuż swojej długiej osi, jak prosiak nad ogniskiem, aby zapewnić równomierny dostęp do powietrza, składników odżywczych i ciągłą nieznaną siłę „ścinającą”, która sprawiła, że komórki przyjęły ostatecznie kształt wydłużony.

Operacja zakończyła się sukcesem. Oskrzele przyjęło się i pokryło zupełnie normalnym, zdrowym śluzem, bez śladów stanu zapalnego. 10 dni po operacji pacjentkę wypisano ze szpitala, a jej stan zdrowia uległ stałej poprawie: była w stanie robić rzeczy wcześniej dla niej niemożliwe, jak wchodzić po schodach bez zadyszki i swobodnie zajmować się swoimi dziećmi. W pierwotnym artykule opisano badania kontrolne na przestrzeni 3 miesięcy, a w 2013 roku opublikowano podsumowanie tego przypadku z perspektywy 5 lat: przeszczep wciąż funkcjonował w stu procentach prawidłowo.



Pusty organ czeka na komórki

Z każdym kolejnym rokiem udaje się odtwarzać coraz bardziej złożone elementy ciała ludzkiego. Świętym Graalem są oczywiście kompletne organy, jak nerka, wątroba albo płuco. Póki co nie udało się jeszcze wyhodować „od zera” żadnego z nich, choć w ostatnich kilku latach pojawiły się obiecujące sygnały świadczące o możliwości stworzenia pełnych rusztowań dla tych organów.

W marcu 2020 roku opisano udaną decelularyzację całej ludzkiej wątroby^[101], a w marcu 2021 roku całego świńskiego płuca^[102]. W obu przypadkach jednym z kluczowych składników „koktajlu” wymywającego komórki był Triton X-100, popularny detergent mający tysięczne zastosowania przemysłowe, medyczne i domowe. Gdybyście się zastanawiali, to całkowicie wypłukana z komórek wątroba ludzka wygląda jak zwykła wątroba, jednak zupełnie pozbawiona swojej charakterystycznej barwy – jest bladoróżowa i półprzezroczysta. Autorzy obu prac zarzekają się, że otrzymane tkanki są w stu procentach niezniszczone i gotowe na przyjęcie komórek.

Cytowane artykuły opisywały wyłącznie procedurę udanego „odkomórkowienia”, bez próby zaludnienia komórkami docelowymi; na tym froncie postęp jest nieco wolniejszy i nawet zaludnienie małej porcji tkanki uznawane jest za sukces. W 2017 roku, przykładowo, opisano „recelularyzację”, czyli ponowne zasiedlenie komórkami, jednego z płatów wątroby świńskiej^[103]. Posłużono się mieszaniną trzech rodzajów komórek: 60 milionów hepatoblastów (komórek progenitorowych wątroby, które w sprzyjających warunkach mogą przemienić się w niektóre dojrzałe komórki tego organu), 45 milionów komórek nabłonkowych z krwi pępowinowej i 12 milionów komórek macierzystych ze szpiku kostnego. Zawiesinę taką cierpliwie przepompowywano przez cały organ, po prostu podpinając się pod naczynia krwionośne prowadzące do wątroby i z niej wychodzące, przepuszczając płyn „w kółko” przez tydzień, co drugi dzień wymieniając go na nową porcję ze świeżo wyhodowanymi komórkami. Autorzy zaskakująco skąpo opisują powodzenie całej procedury, trudno więc powiedzieć, na ile była ona skuteczna. Nie ma żadnych informacji o funkcjonalności tak powstałego „organu”.

Korzystanie ze struktury rzeczywistego, „żywego” organu ma oczywiste zalety – nie trzeba pracowicie rzeźbić tkanki w skali mikroskopowej i w pewnym sensie „gwarantowana” jest prawidłowa struktura przestrzenna docelowego organu. Nie zawsze jednak można sobie pozwolić na taki luksus, a hodowanie zwierząt na organy budzi

wątpliwości etyczne. Najbardziej „czystym” i eleganckim rozwiązaniem jest drukowanie rusztowania, a być może nawet i całego organu, wyłącznie przy pomocy surowych materiałów i własnych komórek pacjenta. Istnieją dziś już dziesiątki „biotuszy” stosowanych przy tworzeniu rusztowań dla komórek, zarówno naturalnych (na przykład agaroz, celuloza i kolagen), jak i syntetycznych (na przykład politlenek etylenu albo polilaktyd). Jeżeli zaś chodzi o część komórkową, to stosowane mogą być dowolne komórki, które potrafią znieść trudy przeciskania się przez dysze drukarki – dojrzałe czy macierzyste. Bywa też, że drukuje się od razu przy pomocy wyhodowanych z góry większych podjednostek, jak pęcherzyki zbudowane z komórek (zob. ramka poniżej).

Jednym z największych nierozwiązanych problemów inżynierii tkankowej jest **odtworzenie sieci naczyń krwionośnych**, które muszą wszak docierać w każde miejsce w organie. Komórki są niezwykle delikatne i fatalnie reagują na ograniczenie ich dostępu do krwi. Objawy niedotlenienia i niedożywienia komórki mogą się pojawić już wtedy, gdy jej odległość od najbliższego naczynia krwionośnego wynosi więcej niż 200 μm (200 mikrometrów, czyli 0,2 mm). Oszacowano kiedyś^[104], że łączna długość wszystkich naczyń włosowatych (albo „kapilar” – tak określa się najcieńsze naczynka układu krwionośnego) w jednym tylko płucu wynosi 790 km. Aby wydrukować funkcjonalne płuco, należałoby odtworzyć tę niewyobrażalnie złożoną sieć. Są badacze, którzy uparcie próbują korzystać z drukarek, jednak tworzenie pustych w środku, drożnych rurek o średnicy liczonej w mikrometrach nie jest wykonalne metodą wprost. Ciekawym pomysłem jest drukowanie samego szkieletu naczyń krwionośnych z jednego materiału, następnie „obtaczanie go” komórkami nabłonkowymi, a ostatecznie chemiczne rozpuszczanie szkieletu – już w 2007 roku udało się^[105] wykonać taką metodą połączoną sieć naczyń o wewnętrznej średnicy 6 μm .

Inną strategią jest wymuszenie powstawania naczyń krwionośnych w wydrukowanym organie przy pomocy metod, mówiąc ogólnie, „biologicznych”. Bądź co bądź, ludzka sieć naczyń krwionośnych powstała samodzielnie w toku rozwoju embrionalnego i nawet u dorosłego człowieka posiada wrodzoną zdolność rozbudowy. Problemem tym żywo zainteresowali się między innymi onkolodzy, którzy z niepokojem śledzą naczynia krwionośne wnikające w guz nowotworowy. Inżynierowie tkankowi i badacze raka równie usilnie próbują więc zrozumieć, jakie są mechanizmy odpowiedzialne za unaczynienie (technicznie: „waskularyzację”) tkanek.

Jeszcze innym kątem natarcia na ten problem jest badanie tzw. samoorganizacji: jest to ogólna nazwa na zdolność układów naturalnych do spontanicznego organizowania się przestrzennego. Mimo wszystko, duża rzeka wygląda na mapie dość podobnie do układu krwionośnego: kilka głównych dopływów, a potem setki i tysiące coraz to mniejszych strumyczków. W podobnie finezyjny sposób potrafią się też organizować kryształki lodu rosnące na szybie. Biotechnolodzy spotykają się więc z fizykami, aby wspólnie pracować nad wymuszeniem w ludzkiej tkance tego typu spontanicznych zjawisk. Rozwiązanie problemu unaczynienia tkanek prawdopodobnie narodzi się gdzieś na styku wszystkich wymienionych wyżej badań.



Mysia tarczycza z drukarki 3D^[106]

Artykuł z 2017 roku. Autorzy postanowili odtworzyć myszą tarczycę – względnie prosty organ, przynajmniej w porównaniu z płucem czy wątrobą. Zasadniczą jednostką organizacyjną tarczycy w skali mikroskopowej jest pęcherzyk tarczycowy: bąbełek otoczony warstwą komórek nabłonkowych. Bąbelki takie poprzątkane są gęstą siecią naczyń krwionośnych, które odpowiadają za przejmowanie hormonów produkowanych przez ten organ. Tarczycza jest więc idealnym poligonem badawczym dla inżynierów tkankowych: z jednej strony ma względnie nieskomplikowaną budowę, a z drugiej jej prawidłowe funkcjonowanie wymaga stworzenia sieci naczyń krwionośnych.

Sztuczna tarczycza opisana w tym artykule została skomponowana z dwóch rodzajów pęcherzyków, naniesionych drukarką 3D na rusztowanie wydrukowane z kolagenu. Świeżo wydrukowany organ przypominał nieco porcję kawioru, w którym poszczególne kuleczki miały jednak rozmiar mikroskopowy. Były to:

- pęcherzyki tarczycowe, pobrane z mysich embrionów, a więc niedojrzałe, ale zupełnie funkcjonalne „cegiełki” budujące tarczycę – istnieje też możliwość tworzenia takich pęcherzyków „od zera”, jednak w tym przypadku postanowiono nieco „pójść na skróty”;
- pęcherzyki utworzone z tkanki omocznia; tu wyjaśnienie: omocznia to jedna z błon otaczających wzrastający embrion (czy to ludzki, czy myszy), w której dochodzi do intensywnego powstawania naczyń krwionośnych, później odżywiających płód.

Autorzy już wcześniej sprawdzili, że gdy tkankę omocznia uformuje się w kształt mikroskopijnych bąbelków, to zachowują się one jak „zarodki” naczyń krwionośnych. Okazuje się, że gdy przyciśnie się je do siebie i hoduje przez pewien czas w swoim

sąsiedztwie, to mają tendencję do zlewania się ze sobą – trochę jak przyciśnięte do siebie bańki mydlane albo oka na rosole. Gdy więc hodowano „kawior” zbudowany z mieszaniny pęcherzyków tarczycowych i omocznionych, po pewnym czasie te drugie stworzyły spontanicznie sieć naczyń krwionośnych, w której rozproszone były w pełni funkcjonalne pojedynczo tarczycy.

Ba, okazało się, że wyhodowany w ten sposób organ można wszczepić myszom z niedoczynnością tarczycy i jest on w stanie prawidłowo funkcjonować, m.in. przywracając właściwą regulację temperatury ciała i zapewniając odpowiedni poziom tyroksyny, ważnego hormonu tarczycowego. Ciekawostka: drukowane tarczycy wszczepiono myszom nie w szyję, gdzie organ ten znajdował się pierwotnie, lecz przytulono je do nerki. Dobrze ukrwiona tkanka, znajdująca tuż pod tzw. otoczką nerki, świetnie bowiem nadaje się jako „gniazdo” dla tarczycy.

Ostatnia uwaga w powyższej ramce mogła przypomnieć nam, że przy naprawianiu człowieka wcale nie trzeba się trzymać kurczowo jego wrodzonej anatomii. Jeśli odpowiednim miejscem dla nowej tarczycy jest bezpośrednie sąsiedztwo nerki, to niech tak będzie. Można oczywiście iść dalej: czy naprawdę musimy pieczołowicie odtwarzać tarczycę według jej zwykłego schematu anatomicznego, zrazik po zraziku i pęcherzyk po pęcherzyku? Bądź co bądź, sztuczne serce ma tylko spełniać funkcję serca, a podobieństwo anatomiczne to kwestia zupełnie drugorzędna – dostępne dziś na rynku sztuczne serca cechują się więc tylko zgrubnym podobieństwem do prawdziwych organów. Czy cała ta inżynieria tkankowa nie jest aby ślepą uliczką?

Najprostszą odpowiedzią na to pytanie jest odwołanie się – znowu! – do wielotorowości medycyny, która zмага się z tysiącami odrębnych problemów. Nie ma rozwiązań uniwersalnych, a każda ścieżka badawcza może znaleźć swoje zastosowanie. Przede wszystkim, nie zawsze konieczna jest wymiana całego organu. Jednym z dużych tematów we współczesnych publikacjach na temat inżynierii tkankowej jest regeneracja fragmentów mięśnia sercowego, który uległ uszkodzeniu w wyniku niedokrwienia – w takich przypadkach potrzebny jest po prostu „plasterek” tkanki serca, którą organizm przyjmie „jak swoją”.

Co jednak z całymi organami? Niewykluczone, że w przyszłości wszczepiane będą zarówno takie, wykonane z materiałów niebiologicznych, jak i tworzone metodami inżynierii tkankowej. Trudno jednak wyobrazić sobie, aby naśladowanie naturalnych tkanek ludzkich naprawdę miało się kiedykolwiek stać „ślepą uliczką”. Nosimy w sobie struktury, które powstały na drodze dziesiątków, a nawet setek milionów lat ewolucji, co czasem niełatwo przebić.

Próba „wymyślenia płuca na nowo” może ostatecznie sprowadzić inżynierów na dawno już wydeptane ścieżki. W cytowanym wyżej artykule, w którym oszacowano długość naczyń krwionośnych w płucu ludzkim, stwierdzono też, że przy objętości 2 litrów ma ono powierzchnię ponad 40 metrów kwadratowych – to tak, jak gdyby całą podłogę niewielkiego mieszkania upchnąć we wnętrzu dwulitrowej butelki wody. Jest to w dodatku struktura lekka i elastyczna. Ewidentnie kryją się za tym miliony lat procesów optymalizacyjnych. Aż kusi, żeby sięgać do rozwiązań sprawdzonych biologicznie.

Co więcej, tkanki i organy dobrze udające te naturalne i zbudowane z komórek ludzkich mają sporo zalet, już choćby przez to, że mają od razu odpowiednią wielkość, kształt, ciężar i konsystencję. Sztuczne płuca, które byłyby na przykład cięższe albo wyraźnie mniej elastyczne od naturalnych, mogłyby prowadzić do problemów czysto mechanicznych. Co więcej, organy „identyczne z naturalnymi” traktowane przez organizm jako „własne” mogą podlegać naturalnym mechanizmom naprawczym: plastikowa rurka zanurzona w tkankach ludzkich nie będzie regenerowana na bieżąco, tymczasem nowa tchawica Claudii Castillo – tak. Dalej: w miarę wzrostu naszej wiedzy o anatomii i fizjologii często okazuje się, że poszczególne organy mają więcej funkcji niż mogłoby wynikać z ich słownikowej definicji. Płuco, przykładowo, pełni ważną rolę w układzie odpornościowym. Jelita są tak silnie unerwione, że niektórzy mówią wręcz o drugim mózgu w naszych brzuchach. Mało tego, atlasy ludzkich komórek i tkanek wciąż zawierają białe plamy i nikt tak naprawdę nie wie, co dokładnie kryje się w mięszu naszych wątrób i trzustek. Jeśli natomiast pozwolimy naszym własnym komórkom macierzystym samodzielnie zorganizować się w odpowiednie tkanki, to można liczyć na to, że przy okazji powstaną również te ich elementy, o których istnieniu możemy nawet nie wiedzieć.

Krótko mówiąc, inżynieria tkankowa nieprędko odejdzie od programu doskonałego naśladowania naturalnych organów.

Odłóżmy może jednak na bok ideę wymiany tkanek i organów na nowe i spróbujmy podejść do problemu zupełnie inaczej. A co, gdyby sprawić, aby te stare, które dostaliśmy w „pakiecie urodzeniowym”, nigdy się nie zepsuły?

IV.1.2. Nanocząstki i nanoroboty

„Nanotechnologia” to kolejne słowo wytrych, które obejmuje wszystkie technologie „działające się” w skali nano-, czyli miliardowych części metra^[67*]. Kiedy producenci paneli słonecznych albo katalizatorów „rzeźbią” powierzchnie swoich podzespołów w skali nanometra – to jest to nanotechnologia. Elektronika już dawno zesłała poniżej skali mikrometra, a więc produkcja układów scalonych to też nanotechnologia. W kontekście biologii i medycyny istotne jest to, że skala „nano” obejmuje obiekty mniejsze od typowej komórki ludzkiej: nanocząstka o wymiarach, powiedzmy 100 nm, mogłaby już wędrować po wnętrzu komórki, a nanorobot o średnicy nawet 1000 nm (czyli 1 μm) wpuszczony do układu krwionośnego albo limfatycznego swobodnie przemierzałby nasz organizm, mieszcząc się nawet w najcieńszych naczyniach włosowatych i mogąc się przeciskać między komórkami. Dużo głębiej niż skala „nano” nie da się już zejść^[68*]: pojedyncze atomy mają średnice zbliżone do jednej dziesiątej nanometra (czyli angstroma, Å): na kwadraciku o boku jednego nanometra można by więc położyć kilkadziesiąt osobnych atomów albo kilkanaście niewielkich, kilkuatomowych cząsteczek.

Znów stajemy przed problemem kłęski obfitości – po kilku wieczorach spędzonych przy najnowszych wydaniach czasopism fachowych i artykułach przeglądowych na temat „nanobiotechnologii” kręci mi się w głowie z nadmiaru pomysłów. Ten rozdział będzie więc chyba jednym z bardziej chaotycznych w całej książce: opowiem w nim po prostu o kilku interesujących projektach, które opisano w ostatnich latach, bez większej nadziei, że złożą się ona w jakąś jedną wizję, jaka wynika choćby z poprzedniego rozdziału.

Docelowo wszyscy marzymy pewnie o tym samym: o roju nanorobotów krążących po naszym organizmie i rozwiązujących na bieżąco wszelkie problemy medyczne, zapewniając nam zdrowie i długowieczność. Cóż, nie mogę obiecać, że kilka poniższych doniesień zadowolą tych, którzy mają nadzieję, że rój taki jest tuż za rogiem. Ale same nanoroboty mogę już obiecać z czystym sumieniem.

Zacznijmy jeszcze od terminologii. Sporo będę pisał o **nanocząstkach**. To ogólne określenie na cząstki o rozmiarach 1–100 nanometrów^[107]. Patrząc „od dołu”, nanocząstki są bliskimi sąsiadami zwykłych cząsteczek chemicznych („molekuł”): większe cząsteczki, jak białka, mają już rozmiary sięgające dziesiątek nanometrów. O zwykłej, pojedynczej cząsteczce chemicznej występującej w organizmie – albo o drobnym pyłku skalnym o rozmiarach nanometrycznych – nie mówi się jednak zwyczajowo „nanocząstka”. Choć nie ma żadnych

oficjalnych definicji, które by to regulowały, to przyjęło się, że terminem tym określa się obiekty zaprojektowane i wytworzone w jakimś określonym celu – a więc cząstki „wycinane”, a nie po prostu występujące naturalnie. Jest w tym trochę marketingu. Prosty przykładem są nanocząstki srebra, mające właściwości antybakteryjne, przeciwwirusowe i przeciwgrzybiczne, a będące w istocie drobnymi kryształkami srebra. Patrząc zaś „od góry”, szczególnie duże drobinki tego typu stają się w pewnym momencie „po prostu” porcjami danego materiału: bardzo duża nanocząstka srebra staje się po prostu „kawałkiem srebra”.

Ciekawa jest natomiast kwestia różnicy między nanocząstkami a **nanorobotami** – to drugie słowo brzmi już naprawdę dumnie, tymczasem ostrej różnicy między nimi nie ma. W praktyce tendencja jest taka, że te „sprytniejsze” albo bardziej ruchliwe nanocząstki – na przykład posiadające zawias albo mechanizm napędowy – określa się jako nanoroboty. Zresztą, dosyć tego teoretyzowania, czas na jakiś przykład.



Nanokapsułka z lekiem reagująca na 5 bodźców^[108]

Jednym z dużych tematów w medycynie jest „sprytne” uwalnianie leków – wspomniałem już o tym pobieżnie w ramce na stronie 81 przy okazji opowiadania o technologiach druku 3D. W większości przypadków substancja lecznicza jest tak naprawdę potrzebna tylko w jednym miejscu, a wiele skutków ubocznych wynika z niepotrzebnego uaktywniania się leków tam, gdzie nie jest to potrzebne. Ba, pomyślmy o dramatycznym przypadku walki z chorobą nowotworową. Popularny model chemioterapii opiera się na substancjach uniemożliwiających komórkom podział, a więc i wstrzymujących powstawanie w organizmie nowych komórek. Idea jest taka, żeby guz przestał się rozrastać – komórki nowotworowe wręcz *definiuje się* jako takie, które „niepotrzebnie” się dzielą. Skutkiem ubocznym jest jednak utrudnianie podziału również i tym komórkom, które dzielą się „śluszenie”, na przykład w skórze (stąd wypadanie włosów) albo w wyściółce jelita (stąd problemy z układem pokarmowym). Czyż nie byłoby pięknie, gdyby cząsteczki środka chemioterapeutycznego „wiedziały”, gdzie mają się aktywować? Idealny scenariusz jest więc taki, że cząsteczkę leku umieszczamy w „sprytnej kapsułce”, która uwalnia go tylko w ściśle określonym miejscu, na przykład na powierzchni lub we wnętrzu guza nowotworowego, ale też na przykład tylko w jednym organie.

Wnętrze guza nowotworowego może różnić się od zwykłej tkanki na wiele sposobów, m.in. ze względu na stężenie tlenu, kwasowość albo obecność określonych związków chemicznych. Od lat prowadzi się próby umieszczenia cząstek leku w „nanokłatkach” albo „nanobąbelkach”, które otwierają się tylko w środowisku, powiedzmy, niskotlenowym. Z czasem opracowano nanocząstki reagujące na coraz więcej bodźców, aby możliwe było jak najdokładniejsze ich „zaprogramowanie”. W przeglądzie na ten temat z 2019 roku^[109] jako rekordzistę wymieniono nanocząstkę wyposażoną w aż 5 tego typu „czujników” – sięgnijmy więc do artykułu źródłowego, żeby przekonać się, jak wyglądają dzisiejsze „sprytne” nanocząstki.

Opis ich produkcji to typowy chemiczny koszmarek. O, proszę, oto jeden z wielu kroków syntezy: „28,1 milimola 5-hydroksy-2-nitrobenzaldehydu dodano do 50 ml etanolu. Następnie powoli dodano 44,96 milimola borowodoru sodu w temperaturze 0°C. Gdy bąbelki niemal zniknęły, mieszaninę mieszano w 25°C przez 24 godziny”^[69]. Uczuła nas to na fakt, że nanocząstki te są raczej „bardzo skomplikowanymi i dużymi cząstkami chemicznymi”, tworzonymi przez chemików, a nie „bardzo prostymi i małymi maszynkami”, tworzonymi przez inżynierów-robotyków.

Procedura ta prowadzi ostatecznie do powstania maleńkich kuleczek (technicznie: „miceli”) wykonanych z włóknistego polimeru, które zachowują się trochę jak kłębki nieznacznie lepkiej włóczki. Wymieńmy teraz kolejno 5 bodźców, na które reagowały owe kłębuszki. Ponieważ polimer ów zbudowany był tak naprawdę z trzech różnych składników chemicznych^[70], sam w sobie reagował już na rozmaite bodźce: kurczył się w wysokiej temperaturze (1) oraz puchł w środowisku kwaśnym (2) i w obecności nadtlenku wodoru (3), czyli wody utlenionej (H_2O_2). Następnie sąsiadujące ze sobą nitki połączone dodatkowymi chemicznymi „mostkami”, które wykazywały swoją własną aktywność w obecności związku o nazwie DTT (4), mającego tysięczne zastosowania w medycynie, również w onkologii. Przygotowane w ten sposób nanocząstki reagowały ponadto puchnięciem na promieniowanie ultrafioletowe (5).

Ostatecznie powstały więc postrzępione kuleczki o średnicy około 120 nanometrów, reagujące na 5 bodźców, w środku których można zanurzyć dowolną substancję, leczniczą lub trującą. Autorzy testowali swoje kuleczki przy pomocy nieszkodliwego barwnika o nazwie czerwien nilu (*Nile red*). Puchnięcie nanocząstek powodowało uwalnianie się barwnika, ukrytego między „nitkami” polimeru, przy czym różne bodźce i ich kombinacje uwalniały go w różnym stopniu i tempie. Przykładowo, w środowisku kwaśnym nanocząstki dość szybko „wypuściły” 31% molekuł barwnika, zaś w obecności nadtlenku wodoru uwolnione zostało 50% barwnika, jednak na przestrzeni 6 godzin. Tempo i stopień wydostawania się barwnika można też było podnosić poprzez dodatkowe naświetlanie promieniowaniem UV. Uzyskano więc nanocząstki, które mogą nie tylko aktywować się w ściśle określonych warunkach, ale których „skuteczność” można miejscowo podnieść, na przykład poprzez naświetlanie lampą UV.

Nie trzeba wielkiej wyobraźni, aby ujrzeć w opisanej wyżej nanocząstce „robota” – wystarczy zapomnieć na chwilę o jej anatomii i skupić się na jej aktywności. Bądź co bądź, stworzono obiekt zdolny do wyrzucania z siebie ładunku w kontrolowany sposób w reakcji na ściśle ustalone bodźce. Gdyby skonstruować urządzenie spełniające dokładnie taką samą funkcję, jednak mające rozmiar kubka do kawy, czujnik światła i strzykawki, nikt nie zawahałby się ani sekundy, aby

nazwać je robotem. W rzeczywistej praktyce nanotechnologii nikt nie wyklóca się jednak o nazwy: naukowcy i inżynierowie są zbyt zajęci faktycznym tworzeniem tych cudeniek, żeby bawić się w cyzelowanie definicji. Zwykle jednak słowo „robot” zarezerwowane jest dla struktur mających ruchome części, a więc choć trochę przypominających zwykle maszyny. To może tłumaczyć, dlaczego opisane wyżej „urządzenie” zaanonsowano w tytule artykułu jako nanocząstkę, a następny bohater tego rozdziału jest już reklamowany jako „nanorobot”. Zacznijmy jednak od omówienia technologii, przy pomocy której go skonstruowano.



Origami DNA

Ideę „origami DNA” opisał w 2006 roku chemik Paul Rothemund i do tego czasu pozostaje ona absolutnym hitem w świecie nanotechnologii. Kiedy czyta się współczesne artykuły na temat nanorobotów medycznych, z grubsza połowa z nich nawiązuje do tej technologii. Samo określenie jest nieco mylące: owszem, obejmuje ono konstruowanie obiektów w skali nano z DNA, jednak nie dzieje się to poprzez składanie arkusza, jak w prawdziwym japońskim origami.

Zwykle myślimy o DNA głównie jako o nośniku informacji genetycznej, jest ono tymczasem również bardzo praktycznym materiałem konstrukcyjnym. Nanotechnolodzy zaś stale poszukują materiałów, przy pomocy których mogliby tworzyć swoje „nanomaszynki”. Materiały pochodzenia biologicznego mają już „na start” przynajmniej dwie wielkie zalety. Po pierwsze, po dekadach badań z zakresu biologii molekularnej znamy całkiem nieźle ich rzeczywiste możliwości: o ile musimy się domyślać, jak wspinać rzeczy można by „utkać” z nanorurek węglowych albo pozlepiać z nanocząstek srebra, wystarczy zajrzeć do podręczników biologii komórki, aby ujrzyć funkcjonalne „nanomaszynki” zbudowane z aminokwasów albo kwasów nukleinowych. Po drugie, mamy niejako zagwarantowaną ich zgodność z tkankami żywymi, czyli tzw. biokompatybilność: organizm żywy nie będzie próbował wydalić albo otorbić nanorobota wykonanego z materiałów, z których on sam jest zbudowany, czego nie można powiedzieć na przykład o nanorobotach metalicznych.

DNA w komórkach ludzkich występuje niemal wyłącznie jako podwójna nić (słynna „podwójna helisa” lub „podwójna spirala”), jednak jej prostszą postacią, występującą choćby w wielu wirusach, jest nić pojedyncza^[71]. Szybka powtórka: nić DNA składa

się z „koralików” należących do 4 różnych odmian, nazywanych literami A, C, G i T. Podjednostki A i T oraz C i G są względem siebie „komplementarne” („zgodne”, „odpowiednie”), co oznacza, że koralik A i koralik T czują do siebie delikatne przyciąganie, jak gdyby były drobnymi magnesikami – jednak żaden z nich nie czuje tego typu przyciągania do koralików C albo G. Z drugiej strony podjednostki C i G mają tendencję do sklejania się ze sobą nawzajem. Dzięki temu możliwe jest bardzo precyzyjne sklejanie ze sobą wielu krótkich nici DNA – albo fragmentów dłuższych nici.

Dzisiaj istnieją już możliwości syntetyzowania nici DNA o dowolnej sekwencji (kolejności) podjednostek. Mogłbym więc wytworzyć „nitkę” o sekwencji, przykładowo, AAAAAAAAAA – składającą się z dziesięciu połączonych ze sobą podjednostek A. Przypuśćmy, że tworzę dokładnie milion takich nitek, a ponadto też, niezależnie od siebie, po milionie nitek CCCCCCCCC, GGGGGGGGGG i TTTTTTTTTT. Następnie zlewam wszystkie cztery próbówki do jednego naczynia. Co się stanie? Cóż, nitki typu „A” skleją się z nitkami typu „T”, a nitki „G” z nitkami typu „C”, ostatecznie z 4 milionów pojedynczych nitek powstaną 2 miliony „podwójnych helis” – milion o budowie „AT” i milion o budowie „CG”.

Idźmy dalej. Przypuśćmy, że tworzę nitkę o budowie AAAAAAAAAAGCGCGCGCTTTTTTTTTT. Co teraz? Gdybym pozostawił ją, pojedynczo, samą sobie, prawdopodobnie „złożyłaby się” sama na siebie: końcówka „A” zlepiałaby się z końcówką „T”, a pośrodku powstałaby pewnie pętelka. Gdybym wprowadził do naczynia milion takich nitek, powstałby ciekawy kogel-mogel zbudowany z rozmaitych konfiguracji. Przykładowo, końcówka „A” jednej nitki mogłaby się „skleić” z końcówką „T” innej nitki.

Pomyślmy teraz o dokładnie tej samej logice zastosowanej do pojedynczej, bardzo długiej nici DNA, która skleja się sama ze sobą w ściśle określonych punktach „zaczepu” – tak sprytnie przemyślanych, żeby był tylko jeden sposób na jej zlepienie się. Nitka taka, jeśli pozostawić jej trochę czasu, powinna swobodnie „poskładać się” w pewien określony kształt – teraz jest chyba jasne, skąd skojarzenie z origami? Taki był właśnie pierwotny pomysł Rothemunda, do dzisiaj będący podstawą technologii origami DNA. Metoda jest genialna w swojej prostocie, a dzisiaj możliwe jest już wymyślenie „na papierze” niemal dowolnej struktury dwu- lub trójwymiarowej, a następnie zrealizowanie jej w skali nano dzięki wyprodukowaniu nitki DNA o ściśle określonej sekwencji. Nitka ta, pozostawiona sama sobie, spontanicznie zwinnie się w odpowiedni sposób – i tyle. Potrzeba oczywiście sporej pomysłowości i ostrożności, aby nie doszło do niepożądanego sklejania się ze sobą sąsiednich nitek.

Naukowcy szybko nauczyli się tworzyć przy pomocy DNA dowolne płaskie, a w 2011 roku już praktycznie dowolne przestrzenne struktury, nawet o bardzo złożonym kształcie. Z czasem wyłonił się też standardowy dziś model, w którym pojedynczą długą główną nić (tzw. „rusztowanie”, *scaffold*) związa się w kształt docelowy przy pomocy „zszywek” (*staples*) – krótkich fragmentów DNA, których funkcją jest wytwarzanie połączeń między ściśle określonymi miejscami na nici głównej. Możliwe jest też oczywiście konstruowanie origami z wielu niteczek.

Cała sztuka projektowania origami DNA polega na wymyśleniu takiej sekwencji rusztowania i takiego zestawu zszywek, żeby całość mogła się zwinąć tylko na jeden sposób. Już w 2009 roku udostępniony został stosowany do dzisiaj program caDNano, który ułatwia projektowanie nici, a w 2016 roku opisano program, który automatycznie generuje sekwencję cząstki DNA po zadaniu mu docelowego kształtu, w jaki ma się ona zwinąć^[110]. Idea była taka, aby użytkownik po prostu „rysował” dowolnie wymyślony kształt 3D, na przykład model swojego nanorobota, a już zadaniem algorytmu byłoby opracowanie gotowej sekwencji. Potem pozostaje tylko wyprodukować cząstki DNA o takiej sekwencji zwykłymi metodami syntezy

chemicznej – co też dzieje się już dziś półautomatycznie. W praktyce jednak do dziś projektowanie nanostruktur DNA wymaga żmudnej pracy przy optymalizacji projektu, a także sporych zasobów obliczeniowych – przewidywanie docelowej struktury sporego origami DNA to robota dla superkomputera.

Najpiękniejszy w tej technologii jest aspekt samoorganizacji: nanocząstki DNA samodzielnie związają się w odpowiedni kształt po prostu wskutek lokalnych oddziaływań fizycznych i chemicznych. Jeżeli dobrze zaprojektuję moje DNA, to będzie tylko jeden sposób na jego związanie. Przy odrobinie staranności można za jednym zamachem wytworzyć astronomiczne ilości nanostruktur, po prostu „wrzucając” wszystkie składniki do jednej kolby i czekając, aż połączą się ze sobą w odpowiedni sposób. W świeżutkim artykule przeglądowym ze stycznia 2021 roku [\[111\]](#) autorzy piszą już o tworzeniu biliona (10^{12}) [\[112\]](#) nanocząstek DNA w jednym kroku.

Czysto teoretycznie, podobną logikę konstruowania struktur nano można by zrealizować przy pomocy rozmaitych podobnych związków zbudowanych z „komplementarnych” podjednostek, jednak DNA ma jedną dodatkową, gigantyczną zaletę: jest kluczowym składnikiem komórek żywych! Jeżeli więc częścią naszego nanorobota DNA będzie „dyndająca niteczka” zawierająca sekwencję jakiegoś genu [\[72*\]](#), to robot ów przyklei się do ludzkiego genomu tylko w jednym określonym miejscu. DNA da się też łączyć z innymi materiałami; przykładowo, w 2017 roku opisano pierwsze origami DNA-białkowe.

„Origami DNA” to dzisiaj cała wielka gałąź nanotechnologii. Jest dopiero czerwiec, gdy piszę te słowa, a tylko w 2021 roku ukazało się już 1210 artykułów naukowych zawierających frazę „DNA origami”.



Nanorobot DNA walczy z zakrzepami [\[112\]](#)

Poreklamowałem, naobiecowałem – czas na jakiś konkret. Proszę bardzo, „inteligentny nanorobot DNA”. Tak w każdym razie głosi tytuł artykułu z sierpnia 2020 roku.

Jego autorzy postanowili zająć się problemem krzepnięcia. Szybka powtórka: w krwi znajduje się białko o nazwie fibrynogen: ot, jeden z rozpuszczonych w osoczu składników naszej krwi. Gdy jednak sklei się ze sobą wiele cząsteczek fibrynogenu, powstaje „supercząstka” o nazwie fibryna, tworząca nierozpuszczalne włóknienka, m.in. współtworzące skrzep. Reakcję łączenia fibrynogenu w fibrynę przeprowadza enzym o nazwie trombina. Gdy więc we krwi znajduje się za dużo trombiny, powstaje też za dużo włóknienek fibryny, co może być niebezpieczne. W skrajnym przypadku mogą się tworzyć niepożądane skrzepy, zatykające naczynia krwionośne – a więc alarm czerwony. Z drugiej strony, jeśli zaś trombiny brakuje, krzepnięcie krwi spada, co może prowadzić na przykład do niegających się krwotoków. Tak źle i tak

niedobrze, więc we wszystkich przypadkach, gdy czyjś organizm nieprawidłowo kontroluje poziom trombiny, potrzebna jest czuła opieka farmakologiczna.

Tu wkraczają „inteligentne nanoroboty”, jak odważnie stwierdzają autorzy. Przyjrzyjmy się najpierw ich budowie. Rdzeniem nanobota jest otwarty z obu stron cylinder – czyli struktura mająca kształt kubka do kawy bez dna – o średnicy i wysokości około 40 nanometrów, zbudowany całkowicie z DNA (metodą „origami”). Ścianki cylindra są cienkie, dzięki czemu przez jego środek względnie swobodnie przepływa płyn. Wewnętrzne ścianki zostały wyposażone w cztery rzędy cząsteczek, nazwanych literami T, A, O i G, również wykonanych z DNA, pełniących następujące funkcje:

- cząsteczki T: kontrola progu pobudzenia;
- cząsteczki A: czujnik trombiny;
- cząsteczki O: aktywacja blokowania;
- cząsteczki G: blokowanie trombiny.

Zacznijmy od opisanego „logiki” tego nanorobota, na razie bez uwzględnienia cząsteczek T; autorzy produkowali zresztą w celach testowych jego wersje pozbawione tej cząsteczki. No to uwaga.

Cząsteczki A reagują z trombiną znajdującą się w płynie swobodnie obmywającym wnętrze cylindra. Po połączeniu się z trombiną wydzielają dodatkową cząsteczkę (oznaczaną literą I), służącą jako sygnał dla cząsteczek O oznaczających: „blokować!”. Cząsteczki O są jednak tylko posłami – i gdy otrzymają tego typu sygnał w postaci cząsteczki I, „trącają” sąsiednie cząsteczki G, które wykonują już zasadniczą pracę: uwalniają do środowiska kolejną cząsteczkę (o nazwie TA-15), która po spotkaniu się z trombiną przylepia się do niej, uniemożliwiając jej tworzenie fibryny. Tego typu cylinder, zawierający wyłącznie cząsteczki A, O i G, po wpuszczeniu go do krwi działałby więc tak, że w kontakcie z trombiną wypuszczałby ze swojego wnętrza cząsteczki blokujące jej aktywność. Odpowiednio duża dawka takich „nanorobotów” całkowicie dezaktywowałaby więc trombinę, kompletnie wyłączając mechanizm krzepliwości.

To jednak nie koniec – pominięta dotychczas cząsteczka T pełni bowiem kluczową rolę „pokrętła”, którym można dostroić czułość całego urządzenia. Jej zadanie polega na przechwytywaniu niektórych cząsteczek I, jeszcze zanim zostaną „zauważone” przez cząsteczki O. Im więcej cząsteczek T, tym większa liczba cząsteczek I może zostać przechwycona. Dzięki temu, gdy stężenie trombiny we krwi jest niskie, nanorobot *w ogóle* nie blokuje jej aktywności – każda cząsteczka I zostaje przechwycona i żadna cząsteczka O nie otrzymuje sygnału „blokować!”. Kiedy jednak poziom trombiny przekroczy wartość progową, niektóre cząsteczki I przecisną się przez „strażników” w postaci cząsteczek T i wywołają reakcję blokowania. Owo stężenie graniczne można precyzyjnie regulować, dodając lub zmniejszając liczbę cząsteczek T we wnętrzu cylindra. Cylinder kompletnie „zawalony” cząsteczkami T byłby zupełnie нефункциональный: żaden poziom trombiny nie wywoływałby uwalniania cząsteczek TA-15.

Przeprowadzone przez autorów doświadczenia – również w rzeczywistym osoczu krwi ludzkiej – pokazały, że ich nanorobot zachowuje się prawidłowo, w dodatku dzięki zwiększaniu lub zmniejszaniu liczby cząsteczek T (testowano wersje zawierające 0, 4, 9, 13 i 18 sztuk) można naprawdę płynnie regulować „pobudliwość” takiego nanorobota. W wersji pozbawionej kompletnie cząsteczek T nanorobot reaguje od razu na trombinę, próbując zablokować jej aktywność, gdy tylko zostanie wykryta. W wersji nieco „inteligentniejszej” nanorobot „czeka”, aż stężenie trombiny osiągnie pewien określony poziom i dopiero od tego momentu zaczyna działać. Autorom udało się tak „podkręcić” swoje nanoroboty, że „budziły się” dopiero po przekroczeniu poziomu rzeczywiste związane w organizmie człowieka

z powstawaniem skrzepów w krwi, i od tego momentu blokowały trombinę tak długo, aż jej poziom spadł do bezpiecznego poziomu. Wtedy ich aktywność ustawała.

Zauważmy, na jak wiele sposobów można kontrolować sposób działania tych „urządzeń”. Po pierwsze, możliwe jest stworzenie nanorobotów o różnym progu pobudzenia (regulowanym poprzez liczbę cząsteczek T „wetkniętych” w wewnętrzną ściankę cylindra). Co więcej, można też utworzyć mieszaninę nanorobotów, składający się z różnych ich wersji. Zauważmy też, że do organizmu człowieka można wpuścić dowolną liczbę „nanorobotów”, co będzie skutkowało szybszym lub wolniejszym reagowaniem na zmiany stężenia trombiny. Możliwe jest więc skomponowanie indywidualnie dobranego „koktajlu”, który będzie, przykładowo, błyskawicznie (dzięki wysokiemu stężeniu odpowiedniej odmiany) reagował na niebezpiecznie wysokie stężenia trombiny, a dużo łagodniej, w dłuższej skali czasowej, reagował na wahania stężenia w okolicach wartości poprawnej.

Przypuszczam, że przynajmniej część z was, czytających te słowa, powątpiewa w duchu, czy opisane wyżej nanocząstki rzeczywiście można określić jako „inteligentne nanoroboty”. Myślę, że część tego oporu wynika z tego samego zjawiska, które sprawia, że niechętnie określamy algorytmy komputerowe jako „inteligentne” (zob. rozdział 1.2): jeżeli zachowanie jakiegoś wynika ze zrozumiałego, deterministycznego mechanizmu, uznajemy całą rzecz za „głupią maszynkę”. W tym miejscu pojawiają się więc znów te same problemy filozoficzne, z których już trochę się wykręcałem w poprzednich rozdziałach – czy to już „prawdziwa” inteligencja, czy to już „prawdziwy” umysł? Wówczas odwołałem się do pragmatycznego podejścia przedstawicieli „szkoły zadaniowej” – w zupełności wystarczy nam, że coś *zachowuje się* inteligentnie albo że wystarczająco dobrze *udaje* inteligencję.

Czy opisany wyżej nanorobot DNA zachowuje się inteligentnie? Powiem może najostrożniej: jest to obiekt reagujący na bodźce w dający się „zaprogramować” sposób i wykonujący jedną z dwóch form aktywności w zależności od proporcji tych bodźców. Tylko tyle i aż tyle. Naprawdę nie chcę się wyklócać o słówka. Jedno jest pewne: twórcy opisanego tu nanorobota nie owijają w bawełnę, a redaktorzy dość szanowanego czasopisma pozwolili im na to. W artykule przedstawiony jest najzupełniej normalny „diagram logiczny” przedstawiający „algorytm postępowania” nanorobota. Stosowany w tekście język obficie nawiązuje też do języka programistyki i robotyki. Proszę bardzo, oto cytat z końcówki artykułu: „Dzięki prostemu algorytmowi i wydajnemu projektowi modułarnemu, opisany tu system może być wykorzystywany do systematycznego konstruowania nanorobotów do wykonywania innych inteligentnych

zadań terapeutycznych, przy których konieczna jest precyzyjna kontrola nad dawkowaniem”. Możecie też być pewni, że gdy tego typu nanorobot trafi na rynek, będzie reklamowany jako „autonomiczny”, „inteligentny”, a kto wie, czy nie „myślący” robot.

Na koniec tego rozdziału postanowiłem opowiedzieć o zupełnie innym podejściu do problemu nanorobotów. Zaczniemy od pewnej pamiętnej sceny z serialu HBO *Czarnobyl* (2019). Na miejsce katastrofy sprowadzono robota mającego zająć się sprzątaniem wysoce radioaktywnych fragmentów rdzenia powstałych wskutek eksplozji. Robot ten jednak szybko przestał działać, uszkodzony przez mordercze dawki promieniowania. W czasie narady pada dramatyczne pytanie, czy są w ogóle gdzieś na świecie roboty zdolne do wykonania tego zadania. Generał Tarakanow odpowiada: „Nie ma takich robotów”, na co chemik Legasow rzuca cicho: „Bioroboty”, po czym wyjaśnia: „Wykorzystamy ludzi”.

Pod pewnymi względami nic nie jest w stanie zastąpić organizmów żywych, które od miliardów lat doskonałą sztukę przeżycia. Co więcej, niektóre z nich postawiły na miniaturyzację, osiągając rozmiary liczone w setkach nanometrów, będąc jednocześnie wciąż aktywnymi, żywymi, reagującymi na sygnały środowiskowe organizmami. Być może przemknęło już wam przez głowę, że dwie opisane wyżej nanocząstki są mimo wszystko w dużym stopniu pasywne: to tak naprawdę kilkudziesięcionanometrowe porcje materii, skonstruowane z wielką starannością tak, aby w odpowiedzi na określony bodziec zachować się, pozwólmy już sobie na tę ekstrawagancję, „inteligentnie”. Najmniejszym wolno żyjącym organizmem znanym obecnie nauce jest tymczasem *Pelagibacter communis*: bakteria w kształcie pałeczki, która może mieć ledwie 120 nm średnicy i 370 nm długości. Jest to całkowicie samodzielny, wyposażony w normalny genom organizm, który odżywia się, rośnie, dzieli i reaguje na bodźce. A co, gdyby dało się go „zaprogramować” tak, aby wykonywał nasze polecenia? Witamy w świecie **robotyki biohybrydowej**, zwanej też biorobotyką.

Gdyby książka ta miała głównie za cel straszenie, prawdopodobnie poświęciłbym nieco więcej miejsca biorobotyce – to wybitnie niepokojący segment technologii. Osobom o silnych nerwach polecam choćby zapoznanie się z „owadzimi cyborgami” (*insect cyborgs*). Ot, opisany w 2017 roku chrząszcz^[113] z przyklejonym na grzbiecie odbiornikiem i generatorem sygnału podpiętym do elektrycznych stymulatorów mięśni nóg. Żywym owadem wyposażony

w taki zestaw można kierować zupełnie jak zdalnie sterowanym samochodziem: sygnał z „plecaczka” jest silniejszy od sygnału wysyłanego przez mózg chrząszcza, więc po wciśnięciu guziczka „przód” owad po prostu idzie przed siebie. Po zejściu do skali mikro znajdziemy choćby artykuł z 2020 roku^[114], w którym opisano nanorobota skonstruowanego z pojedynczego plemnika ludzkiego zamkniętego w „zbroi” o kształcie miniaturowego pustego pocisku. Pod wpływem pola magnetycznego ów „plemnikobot” (*spermbot*) zaczynał kręcić witką i płynąć w zadanym kierunku.

My jednak skupimy się na zastosowaniach medycznych biorobotyki w skali nano (wielbiciele nowych słów, oto coś dla was: „**nanobiorobotyka**”^[73*]). „Podkradanie” rozwiązań organizmom żywym może następować w najróżniejszej skali. Na poziomie cząsteczkowym może to polegać choćby na posłużeniu się koktajlem naturalnych enzymów, które po zamknięciu w polimerowym „kubeczku”, przypominającym trochę wydrążoną oliwkę o średnicy około 250 nm, zapewniają mu napęd^[115] – powstaje wówczas coś na kształt nanomeduzy, która prze do przodu, gdy z jej otworu „wyrzutowego” wydostają się produkty reakcji.

Na poziomie nieco wyższym można też „pożyczyć” jakiś komponent, na przykład więc bakteryjną do poruszania się. W 2020 roku opisano^[116] nanorobota wyposażonego w „strzykawkę” o długości 300 nm z ruchomą, wysuwaną „igielką” o średnicy zaledwie 40 nanometrów, pierwotnie służącą bakterii *Photorhabdus asymbiotica* do wstrzykiwania toksyn do organizmu ofiary. Po jej zaprogramowaniu tak, aby wbijała się wyłącznie w powierzchnie komórek bakterii chorobotwórczych, stała się swego rodzaju „inteligentnym antybiotykiem”. I znowu to nazewnictwo – czy obiekt, który aktywnie wyszukuje określonego rodzaju bakterie, a następnie przyczepia się do nich, przebija ścianę komórkową i wstrzykuje do środka truciznę, jest jeszcze „lekarstwem”, czy już „robotem”^[74*]?

Można też oczywiście posłużyć się całym gotowym organizmem, który staje się naszym niewolnikiem, a po odpowiednim uzbrojeniu i zaprogramowaniu – „biorobotem”.





Bakteria dostarcza lek do guza nowotworowego [\[117\]](#)

W 2016 roku opisano jeden z nielicznych przypadków faktycznego medycznego zastosowania bakteryjnych biorobotów – jeszcze nie u człowieka, ale już w żywej myszy.

Posłużono się bakteriami z gatunku *Magnetococcus marinus* (nawiasem mówiąc, jest to bliski krewniak wspomnianego wyżej cudu miniaturyzacji, jakim jest *Photorhabdus asymbiotica*), które w warunkach naturalnych są kuleczkami o średnicy około 1000 nm, żyjącymi w słabo natlenionych warstwach wód oceanicznych. Co istotne, magnetokoki posiadają wyjątkową zdolność do wykrywania ziemskiego pola magnetycznego dzięki samodzielnie tworzonym kryształkom magnetytu zamkniętym w specjalnych organellach – magnetosomach. Te dwie własności – reagowanie na pole magnetyczne i upodobanie do niskiego poziomu tlenu – nakierowały badaczy na ciekawy pomysł.

Otóż jedną z bardziej kłopotliwych cech guzów nowotworowych jest występowanie w nich regionów, które nie zdążyły jeszcze wytworzyć własnej sieci naczyń krwionośnych (o problemie z powstawaniem naczyń krwionośnych wspominałem pokrótce w poprzednim rozdziale o inżynierii tkankowej). Znajdujące się w nich komórki są więc trudno dostępne dla wszystkiego, co wstrzykniemy dożylnie – na przykład substancji chemioterapeutycznych. Co gorsze, z nie do końca jasnego powodu takie niedotlenione tkanki nowotworowe słabiej reagują również na radioterapię (naświetlanie). Jak więc je zniszczyć? A może udałoby się „namówić” magnetokoki, aby precyzyjnie się do niedotlenionych regionów guza nowotworowego?

Okazuje się, że metoda ta działa wręcz znakomicie. Naturalne „parcie” magnetokoków ku niskim stężeniom tlenu w połączeniu z lokalizacyjnym działaniem pola magnetycznego (nie chcemy przecież, żeby bakterie rozpełzły się po całym organizmie!) sprawia, że po wstrzyknięciu porcji bakterii w okolice guza nowotworowego, w obecności specjalnie na tę okazję wygenerowanego pola magnetycznego, po niedługim czasie roją się one w bezpośrednim otoczeniu najbardziej kłopotliwych komórek. Autorzy opisywanej tu pracy pokryli powierzchnie bakterii nanobąbelkami (technicznie, były to „liposomy”) o średnicy około 170 nm wypełnionymi substancją o działaniu antynowotworowym o nazwie SN-38.

Aha, drobna uwaga językowa. Tak, autorzy artykułu piszą o „inteligencji” swoich magnetokoków.

Na koniec tego podrozdziału jeszcze tylko krótka uwaga na temat pojęcia, które budzi w ostatnich latach coraz większe emocje: **biologia syntetyczna**. Rozochoceni naukowcy i nieostrożni dziennikarze lubią czasem sugerować, że w laboratoriach powstają dziś „od zera”, tj.

syntetyzowane są, komórki żywe. Nic z tych rzeczy, przynajmniej jeszcze nie. Rzeczywiście, prowadzi się badania pod parasolem „biologii syntetycznej”, mające na celu „podrabianie” niektórych składników organizmów żywych. Ba, w 2016 ogłoszono^[118] pierwszy przypadek wyposażenia bakterii w genom całkowicie stworzony w warunkach laboratoryjnych^[75*]. Warto jednak zrozumieć, na czym polegała ta procedura: zsekwencjonowano genom bakterii o nazwie *Mycoplasma mycoides*, po czym – po „redakcji” polegającej głównie na usunięciu pewnej liczby genów – zsyntetyzowano chemicznie kopię takiego genomu (w ramce na s. 189 wspomniałem, że tworzenie cząsteczek DNA jest już dzisiaj dość dobrze opanowane) i „wklejono” ją do komórki mykoplazmy, z której usunięto jej własny genom. To oczywiście gigantyczny sukces, świadczący ponadto o naszym znakomitym zrozumieniu molekularnych podstaw życia, jednak daleki od „stworzenia życia *in vitro*”.

Ten obszar badawczy będzie oczywiście rozwijany i już za kilka lat będę mógł prawdopodobnie opowiadać o jeszcze bardziej „odchudzonych” wersjach tego typu komórek, zawierających jeszcze precyzyjniej opracowaną listę genów. Czysto teoretycznie, metodą taką można by tworzyć komórki tylko luźno inspirowane realnymi gatunkami bakterii, za to reagujące na zaprojektowane przez nas bodźce oraz trawiące takie pokarmy, produkujące takie cząstki i preferujące takie środowisko życia, jakie tylko sobie wymarzymy. A może udałoby się stworzyć „bakterie” syntetyzujące nanocząstki i nanoroboty, o właściwościach uzależnionych od lokalnych potrzeb występujących, przykładowo, w jelicie albo wnętrzu guza nowotworowego? Jak zwykle, najbardziej szalone wizje wyłaniają się z łączenia kropek.

IV.2. Projektujemy człowieka idealnego

Myszę, że po lekturze poprzedniego rozdziału powinna nam przynajmniej w zarysie malować się wizja „cudownej medycyny przyszłości”, która oferuje hodowanie tkanek i organów z naszych własnych komórek oraz wpuszczonych do naszej krwi rojów nanocząstek, nanorobotów i biorobotów reagujących na bieżąco na wszelkie problemy medyczne i utrzymujących nas w zdrowiu i pięknie.

A co, gdyby problem uciął w zarodku – zupełnie dosłownie? Po co naprawiać w nieskończoność ludzkie ciało – czy nie dałoby się „po

prostu” sprawić, żeby się nie psuło albo przynajmniej psuło znacznie, ale to znacznie rzadziej? Przechodzimy w ten sposób od leczenia do zapobiegania, co w tym kontekście wymaga od nas odważnego zanurzenia się w nasz własny „plan konstrukcyjny” – czyli nasz genom – i naprawienia w nim wszelkich „błędów” i „niedoróbek”. Gotowi?

IV.2.1. Poznaj siebie

Ten rozdział będzie względnie krótki, chociaż – jak zwykle w tej książce – poruszone są w nim naprawdę fascynujące tematy, które trzeba by (i chciałoby się!) omawiać godzinami. W największym skrócie wyjaśnimy jednak może tylko istotę pewnego problemu i wymienimy parę ciekawszych kroków na drodze ku jego przewyciężeniu.

Problem jest naprawdę prosty. Jeżeli naszym celem jest ulepszenie człowieka poprzez edycję jego genomu, jeszcze zanim się narodzi (a im wcześniej, tym lepiej), to musimy najpierw doskonale zrozumieć, w jaki sposób geny przekładają się na budowę i funkcjonowanie ciała. Technicznie, mowa więc o relacji genotyp–fenotyp. Niestety, z każdą kolejną dekadą okazuje się, że jest to jeszcze trudniejsze, niż nam się zdawało ledwie parę lat wcześniej.

Zacznijmy może od przypomnienia, w tempie odrzutowca, co to właściwie są te **geny**. Często stosuję w takich przypadkach metaforę kulinarną. Pomyślmy o genach jako o przepisach, zamkniętych w wielkim zbiorze kulinarnym o nazwie **genom**. Każda komórka mojego ciała^[76*] ma w swoim jądrze komórkowym jedną kopię tego zbioru, zawartego w 46 osobnych księgach o nazwie **chromosomy**^[77*]. Cóż można „ugotować” na bazie tych przepisów? Białka. Po zaniechaniu paru subtelności^[78*] można powiedzieć, że genom zawiera po prostu przepisy na białka. Wszystkie inne elementy żywych komórek, tkanek i organizmów, typu tłuszcze, cukry czy materiały mineralne (jak kość), tworzone są dzięki białkom. Krótko mówiąc, jeżeli mówimy o „genie czegoś”, na przykład „genie niebieskich oczu”, „genie inteligencji” czy „genie otyłości” (to wszystko zmyślone przykłady^[79*]), to ktoś pedantyczny mógłby zapytać, za każdym razem: „OK, no więc o jakim białku mówisz i co ono robi?”, błyskawicznie sprowadzając *wszystkie* cechy organizmu do... białek. Jak to w ogóle możliwe?

W rozdziale IV.1.1 mówiliśmy krótko o rozwoju embrionalnym. Pamiętacie? Była tam mowa o najróżniejszych wpukleniach

i wypustkach oraz grupkach i warstwach tkanek powstających w embrionie pod wpływem rozmaitych bodźców, głównie „sygnałów chemicznych”. Czas chyba wyjaśnić, że owe tajemnicze „sygnały chemiczne” to nic innego jak białka albo tworzone przez nie substancje, należące do sporej klasy tzw. morfogenów. Jest to specjalna grupa związków, produkowanych przez organizm na etapie rozwoju zarodkowego, odczytywanych przez komórki jako sygnał: do przemieszczania się, dzielenia się, transformacji, łączenia się z sąsiadami, a nawet umierania. Jednym z wielkich otwartych problemów biologii jest zrozumienie, w jaki sposób czas i kolejność aktywowania się genów odpowiedzialnych za morfogeny wpływają ostatecznie na naszą anatomię i fizjologię. W ogólności, sekret relacji genotyp–fenotyp w dużym stopniu kryje się w rozwoju embrionalnym człowieka czy też, mówiąc ogólnie, w tzw. **biologii rozwoju**. Dużą pomocą jest oczywiście badanie genetycznych wad rozwojowych: przypadków, gdy pojawia się prowadząca do problemów zdrowotnych mutacja w określonym genie (pomyślmy o niej jako o niefortunnej literówce, powstałej przy kopiowaniu przepisu), brak jakiegoś genu lub genów albo nawet całego chromosomu (lub jego podwojenie).

Pierwszym zadaniem do rozwiązania dla wielbicieli genetycznego udoskonalania człowieka jest zrozumienie genetycznego podłoża określonej cechy. Geny działają zaś przez całe życie, a złożone cechy dorosłych ludzi, jak ich inteligencja, kreatywność, zręczność czy siła wyłaniają się wskutek wieloletnich oddziaływań pomiędzy cechami wrodzonymi i środowiskiem: istny koszmar. Żeby naprawdę uzyskać zdolność genetycznego „projektowania” człowieka, musielibyśmy posiadać pełną wiedzę o historii każdej komórki, każdej tkanki, każdego genu i każdego białka oraz o tym, w jaki sposób złożone cechy wyłaniają się wskutek ich oddziaływań, najpierw z organizmem matki, a następnie środowiskiem. Bagatela!... Musimy więc po prostu poznać całą biologię człowieka, szczegółowo. Dwa ostatnie punkty powyższej listy przypominają nam zaś dobitnie, że nawet najwspanialej zaprojektowany człowiek może trafić na matkę pijącą w ciąży alkohol albo na pielęgniarkę niezgułą, który upuści noworodka na głowę w pierwszej dobie jego życia.

Nie popadajmy jednak w pesymizm i nie próbujmy rozwiązać wszystkich problemów na raz. Jak łatwo się domyślić, naukowcy nie ustają w próbach wydarcia Matce Naturze sekretów embriogenezy i biologii rozwoju. Oto parę doniesień z ostatnich lat, które pokazują skalę naszej ciekawości i postępy w jej zaspokajaniu:

- W 2020 roku opisano^[119] „mapę ekspresji genów” w mysim zarodku. Płody śledzono od 10 dnia ciąży do narodzin (które mają miejsce około tygodnia później). Regularnie pobierano komórki należące do 17 różnych organów i dla każdej z nich sprawdzano, które geny były w danym momencie aktywne. Powstała w ten sposób „ruchoma mapa” aktywności poszczególnych genów z dokładnością do jednego dnia i jednego organu – pozwalająca na przykład sprawdzić, jakie geny aktywowane są w płucu w 13 dniu ciąży.

- W 2018 roku opisano^[120] wczesne etapy rozwoju mysiej trzustki, od dnia 9 do dnia 12, z dokładnością do pojedynczych komórek. (Przy okazji tego badania okazało się, że istnieje nieznaną wcześniej populacja komórek macierzystych trzustki – to tak w nawiązaniu do ramki na s. 173).

- W 2019 roku udało się^[121] opisać rozwój śródmózgowia myszy w dniach 10,5–13,5. Zidentyfikowano 710 genów, których aktywność w tym okresie w tym obszarze mózgu wydaje się mieć największe znaczenie dla przyszłej struktury śródmózgowia.

I tak dalej... i dalej... i dalej...

Zauważyliście może, że żadne z tych badań nie było prowadzone na płodach ludzkich? Cóż, w grę wchodzi tu „subtelna” kwestia etyczna. Od czego jednak ludzka pomysłowość? Chyba już czas na opowieść o jakiejś fajnej technologii.



Organoidy

W rozdziale IV.1.1 była mowa o hodowaniu tkanek ludzkich. Wtedy, jak przypuszczam, nie wzbudziło to wątpliwości etycznych: cóż bowiem moralnie ryzykownego w naczyniu laboratoryjnym z porcją ludzkiej wątroby? Co jednak, jeśli będą to małe embriony? Albo... mózdzek?

„Organoid” to miniorgan, wyhodowany *in vitro*, najczęściej z komórek macierzystych (zob. ramkę na s. 173) wykazujący przynajmniej częściowe

podobieństwo do dojrzałego organu i część jego funkcji. Idea narodziła się w toku pracy z komórkami macierzystymi, które wykazują czasem – o czym przelotnie wspomniałem przy okazji tematu inżynierii tkankowej – zdumiewającą wręcz skłonność do samoorganizacji. Komórki macierzyste nie tylko same zamieniają się w komórki dojrzałe, ale ponadto „dogadują się” ze swoimi sąsiadkami, odtwarzając w pewnym stopniu lokalny porządek właściwy dla odpowiedniej tkanki. Choć nie jest tak dobrze, że porcja komórek macierzystych wątroby nabrana łyżką do lodów i zostawiona w jakimś ciepłym miejscu spontanicznie przemieni się w wątrobę, to ich niewielka grupka, o rozmiarze paru milimetrów sześciennych, może w skali mikroskopowej uporządkować się do postaci prawidłowej tkanki wątroby. Dziś tworzy się już rutynowo organoidy wszelakiego rodzaju: jelitowe, nerkowe, wątrobowe, płucne, językowe, trzustkowe... no i, rzecz jasna, mózgowce – tworzone z komórek progenitorowych odpowiedniego rodzaju.

Osobną kategorię są **organoidy embrionalne**^[80*], tworzone z komórek macierzystych „najgłębszego” typu, odpowiadających komórkom budującym ludzki embrion tuż po jego powstaniu. Ich wytwarzanie i badanie stanowi interesujący problem etyczny: czysto technicznie, nie są to ludzkie embriony – ponieważ nie powstały wskutek zapłodnienia. Nie zachowują się też do końca jak embriony, ponieważ na etapie ich tworzenia nie odtwarza się szczegółowo prawdziwej struktury ludzkiego embrionu – są to mimo wszystko, mówiąc w lekkim uproszczeniu, kolonie komórek. Z każdym kolejnym rokiem udaje się jednak odtworzyć z ich udziałem coraz więcej cech rozwoju embrionalnego – a po to przecież są badane. Przykładowo, w 2019 roku doniesiono^[122], że udało się wstępnie odtworzyć powstanie w ludzkim organoidzie embrionalnym tzw. smugi pierwotnej – charakterystycznej kreski na części grzbietowej embrionu, będącej ważnym punktem odniesienia w jego dalszym rozwoju.



Mózdzek w hodowli zaczyna się budzić

Organoidy mózgowce (*cerebral organoids*) to organoidy składające się z komórek ludzkiego układu nerwowego. Tworzy się je zwykle z indukowanych pluripotentnych komórek macierzystych (zob. ramkę na s. 173), które „wygrzewa się” przez kilka miesięcy, zapewniając im pokarm i rozmaite bodźce do rozwoju.

Jedne z nich mają stanowić miniaturowe wersje całego mózgu – w nich pojawia się wielka różnorodność typów komórek, a czasem również spontanicznie powstają komórki światłoczułe^[123], będące jak gdyby załącznikiem siatkówki oka.

Inne mają stanowić model tylko części mózgu, najczęściej kory mózgowej – to tzw. **organoidy korowe** (*cortical organoids*). Kora wzrokowa to najbardziej zewnętrzna część mózgu. To ona rozwinęła się szczególnie szybko w trakcie ostatnich paru milionów lat ewolucji człowieka, a nasze mózgi są najwyraźniej pofałdowane właśnie po to, aby w ograniczonej objętości naszej czaszki zmieściło się jak najwięcej kory. Nic więc dziwnego, że budzi ona szczególne zainteresowanie twórców organoidów.

Bardzo szybko okazało się, że komórki macierzyste tkanki nerwowej, jeśli „namówi się” je, by przemieniły się w kawałek kory mózgowej, samoistnie odtwarzają jej charakterystyczną 6-warstwową strukturę. Mało tego, po kilku miesiącach rozwoju zaczyna w niej dochodzić do spontanicznej aktywności elektrycznej.

Autorzy pewnego, opublikowanego w 2019 roku, artykułu^[124] przez niemal rok śledzili rozwój swoich organoidów – mających postać kuleczek o średnicy około 1,5 mm i mieszczących w sobie, jak szacują autorzy, mniej więcej jedną milionową liczbę neuronów mieszczących się w prawdziwej ludzkiej korze mózgowej. Szczególnie uważnie badano aktywność elektryczną. Myślę, że najlepiej będzie, jeśli po prostu zamieszczę tłumaczenie fragmentu tego artykułu, poddanego tylko nieznacznej redakcji dla klarowności.

Na wczesnym etapie rozwoju (2 miesiące) okresy ciszy przerywane były krótkimi falami spontanicznej aktywności, obejmującej cały organoid, następującymi mniej więcej co 20 sekund. Fale takie określać będziemy jako „zdarzenie sieciowe”. [...] Od 4 miesiąca pojawił się dodatkowy pik aktywności około 0,3–0,5 sekundy po rozpoczęciu każdego zdarzenia sieciowego, którego zajście inicjowało dodatkową oscylację, o częstotliwości około 2–3 Hz [czyli 2–3 razy na sekundę^[81]]. [...] Regularna aktywność oscylacyjna w czasie zdarzenia sieciowego z czasem przemieniła się w silniejsze, ale i mniej regularne oscylacje. Stopniowo stawały się one coraz bardziej złożone. [...] Dodatkowo, zwiększyła się też przestrzenna i czasowa nieregularność w małej skali (w trakcie zdarzenia), co oznacza, że zachowanie się populacji neuronów od momentu rozpoczęcia się zdarzenia sieciowego przestało być deterministyczne.*

I teraz najlepsze. „Wzorzec naprzemiennych okresów ciszy i zsynchronizowanych fal aktywności przypominał te widoczne w badaniach EEG^[82*] wcześniaków. Od mniej więcej 22–23 tygodnia ciąży okresy braku aktywności przerywane są oscylacjami o wysokiej amplitudzie, trwającymi kilka sekund. W miarę zbliżania się do terminu porodu okresy całkowitej ciszy znikają”. Autorzy przeanalizowali 101 zapisów EEG wcześniaków, w wieku od 24 do 38 tygodni, porównując je z aktywnością swoich organoidów pod kątem 12 ściśle zdefiniowanych kryteriów (typu: odstęp czasowy fal aktywności, czas ich trwania, częstotliwość oscylacji w ich trakcie i tak dalej). Ostatecznie doszli do wniosku, że nie tylko aktywność organoidów korowych przypomina aktywność mózgowców wcześniaków, ale ponadto jedne i drugie mają podobne „ścieżki rozwojowe” – coraz to „starsze” organoidy coraz bardziej przypominały mózgi starszych dzieci.

Dla jasności: osoby pracujące z organoidami mózgowymi doskonale zdają sobie sprawę z ciężaru gatunkowego tego tematu – już choćby dlatego, że są regularnie pytani przez dziennikarzy, czy ich najnowsza kuleczka już próbowała do nich zagadywać i jak się czują, bawiąc się w Boga. Wszyscy odpowiadają jednym głosem, że absolutnie nie ma mowy o wyłonieniu się w organoidach choćby najprostszych postaci świadomości. Madeline Lancaster, jedna z ważniejszych badaczek tych struktur, zapytana, czy jej organoidy są świadome, argumentuje tak^[125]:

- po pierwsze, są na to za małe: 1 milion neuronów to 1/4 tego, co mieści się w korze mózgowej myszy i odpowiada mniej więcej wielkości mózgu raka;

- po drugie, są na to zbyt proste: mózg człowieka składa się z dziesiątków i setek regionów o różnej budowie mikroskopowej, połączonych ze sobą w ściśle określony sposób; organoidy wciąż jeszcze przypominają raczej „kulkę neuronów” niż „mały mózg”;

- po trzecie, czego właściwie miałyby być świadome? – bez narządów zmysłów organoidy nie otrzymują żadnego „materiału do przemyślenia”.

Aby nie wchodzić w filozoficzną debatę z profesorem Lancaster, wyklócanie się z nią pozostawię już wam. Jestem pewien, że macie jej sporo do wytłumaczenia. Ja tymczasem opowiem na koniec tego rozdziału o pewnej sprytniej technologii, która może pozwolić na „ulepienie” z organoidów czegoś w stylu organizmu, bez bawienia się w rekonstruowanie całego ciała.



Ciało na płytce^[126]

Przypuśćmy, że jesteśmy już w posiadaniu zadowalającego organoidu wątrobowego. Co dalej? Aż kusi, żeby sprawdzić, jak zachowałyby się w kontakcie z mózgiem, nerką albo trzustką. Nic prostszego.

Organ na płytce (*organ-on-a-chip*) opisano początkowo w latach 80. jako miniaturową wersję wybranego organu, zrealizowaną poprzez hodowanie kolonii kluczowych dla tego organu typów komórek ludzkich w osobnych naczyniach, połączonych kanalikami. Gdy na rynku biotechnologicznym pojawiły się komórki macierzyste i organoidy, organy na płytkach uznano początkowo za przestarzały sposób robienia czegoś, co teraz można robić lepiej: po co hodować niezależnie od siebie kolonie różnych komórek, skoro można od razu wyhodować organoid, który będzie zawierał wszystkie te typy komórek w jednym miejscu? W ostatnich latach udało się jednak pięknie połączyć obie metody – w poszczególnych przegródkach łądują teraz nie kolonie dorosłych komórek, lecz organoidy. W ten sposób powstaje **ciało na płytce** (*body-on-a-chip*).

W 2020 roku opisano tego typu płytkę zawierającą aż sześć organoidów, reprezentujących: wątrobę, serce, płuco, naczynia krwionośne, mózg i jądro. Każdy z organoidów wyhodowano w osobnej przegródce o średnicy 4 mm, standardowymi metodami znanymi z literatury, z ludzkich komórek macierzystych. Przegródki połączone zostały cienkimi przewodami, którymi pompowano płyn reprezentujący

osocze krwi. Czysto teoretycznie, przegródki można by połączyć w dowolnej konfiguracji, tu jednak postawiono na model „liniowy”, tj. płyn przepływa kolejno przez wszystkie organoidy, w kolejności wskazanej na początku tego akapitu. Autorzy piszą, że pracują nad bardziej realistyczną konfiguracją, jednak nawet teraz każdy organoid unosi się w środowisku chemicznie modyfikowanym przez wszystkie pozostałe. A o to wszak chodzi.

Po co to wszystko? Choćby dla badań toksykologicznych. Autorzy najpierw hodowali swój zestaw przez 14 dni, żeby sprawdzić, czy organoidy „dogadują się” ze sobą. Gdy wszystko wydawało się przebiegać poprawnie, do „żył” wpuszczono ifosfamid – lek onkologiczny, który w organizmie ludzkim ulega aktywacji w wątrobie^[83*]. Zachodzące w wątrobie reakcje nie przebiegają jednak ze stuprocentową skutecznością i część ifosfamidu nie zostaje zamieniona w pożądaną substancję o działaniu przeciwnowotworowym, tylko w związki szkodliwe czy wręcz toksyczne. To więc idealne zadanie dla ciała na płycie! Dzięki obecności organoidu wątrobowego ifosfamid zostaje przekształcony chemicznie jak w prawdziwym organizmie ludzkim (tj. o tyle, o ile jest to „realistyczny” organoid; ten opisywany w cytowanym artykule spał się dobrze), a dzięki obecności pozostałych organoidów jesteśmy w stanie zbadać, jakich można się spodziewać niepożądanych skutków ubocznych po wyprodukowanych w „wątrobie” toksynach. Co więcej, każdy związek chemiczny krążący we krwi może wywoływać rozmaite subtelne reakcje w innych organach, te zaś kolejne, i kolejne. Ostatecznie, na każdy bodziec reaguje po prostu „całe nasze ciało”. Ciało na płycie to idealne laboratorium do badania tego typu złożonych reakcji.

Aha, warto to podkreślić: organoidy można tworzyć również z indukowanych komórek macierzystych, tj. powstałych metodą „cofania zegara biologicznego” dojrzałych komórek dorosłego człowieka (zob. też ramkę na s. 173). Oznacza to, że można stworzyć nie tylko „po prostu ciało na płycie”, ale też „twoje ciało na płycie”. W kwietniu 2021 roku ukazał się już – lekko wizjonerski, ale mimo wszystko naukowy – artykuł właśnie na ten temat, chyba pierwszy taki^[127]. Wylania się z niego wizja, w której lek, którego profil bezpieczeństwa nie jest jeszcze ustalony, a pojawia się potrzeba zastosowania go (na przykład w ramach eksperymentalnej, ratunkowej terapii onkologicznej), testowany jest nie tylko na populacji ludzkiej w toku badań klinicznych, ale ponadto może zostać równolegle przetestowany na ciele na płycie stworzonym przy pomocy tkanek konkretnego pacjenta.

IV.2.2. Edytujemy!

Przypuśćmy, że poznaliśmy już ludzkie ciało i jego rozwój pod wpływem genów na tyle dobrze, że przychodzi nam ochota na eksperymentowanie z modyfikacją jego genów na etapie embrionalnym. Możemy zresztą zacząć od naprawdę prostych przypadków, jak choroby genetyczne o dobrze poznanym podłożu, opierające się na jednej mutacji punktowej w jednym tylko genie. Jakie technologie edycji genomu mamy właściwie do dyspozycji?

Mało jest dziedzin technologii, w przypadku których nastąpił w ciągu ostatniej dekady tak fundamentalny przełom, co w temacie edycji genomu. Jeszcze 10 lat temu na powyższe pytanie trzeba by

odpowiedzieć tak: jasne, da się edytować genom, w tym ludzki, jednak jest to żmudna, kosztowna, wielomiesięczna procedura, o trudnym do przewidzeniu stopniu powodzenia. Dzisiaj odpowiedź ta brzmi zupełnie inaczej: jasne, edytujemy! A wszystko za sprawą CRISPR/Cas.



CRISPR/Cas

CRISPR/Cas to metoda „podkradnięta” bakteriom^[84*]. Okazuje się, że te niepozorne jednokomórkowce wypracowały sobie bardzo sprytny mechanizm zapamiętywania zagrożeń, podobny do tego wykorzystywanego przez nasz własny układ odpornościowy. Niektóre bakterie, zmagając się z infekcją wirusową, wycinają z genomu wirusa małe fragmenty jego kodu genetycznego, a następnie wklejają je w swój własny genom w ściśle określonych miejscach. To podręczna biblioteczka „niebezpiecznych” sygnałów genetycznych, nosząca techniczną nazwę CRISPR. Od tego momentu komórka produkuje stale pewną liczbę białek o nazwie Cas, które należą do rodziny nukleaz, czyli białek „przecinających” kwasy nukleinowe, takie jak DNA. Cały dowcip polega zaś na tym, że nie są to po prostu ślepe nożyczki, rzucające się na oślep na wszystko, co przypomina DNA. Oj, nie. Każde białko Cas sprzęgnięte jest z jedną z próbek „przechwyconego od wroga” materiału genetycznego pochodząca z biblioteki CRISPR. System CRISPR/Cas opiera się więc na rozcinaniu genomu wroga na podstawie przechwyconej wcześniej krótkiej sekwencji „rozpoznawczej”.

Co wewnętrzne sprawy bakterii mają do rewolucji w inżynierii genetycznej? Cóż, genetycy od dekad marzyli właśnie o czymś takim, co w pierwszych latach XXI wieku opisali bakteriologowie – o precyzyjnych „nożyczkach” do DNA, wykonujących nacięcie w ściśle określonym miejscu. Nacięcie to pierwszy krok na drodze do edycji genetycznej. Co istotne, nasze organizmy – jak wszystkie w ogóle w świecie ożywionym – posiadają zdolność do „naprawy” DNA i automatycznie zszywają wszelkie zerwane nitki. Jeżeli więc za sprawą mechanizmu CRISPR/Cas zrobimy w pewnym miejscu nacięcie i odpowiednio szybko dokonamy na miejscu odpowiedniej „podmiany” – może być to coś tak prostego, jak zamiana jednej „cegiełki” DNA (nukleotydu) na inną – to nasze własne komórkowe mechanizmy naprawcze zajmą się „sklejeniem luźnych końców”.

W 2012 roku grupa badawcza, do której należały m.in. późniejsze laureatki Nagrody Nobla, Jennifer Doudna i Emmanuelle Charpentier, jako pierwsza opisała metodę edycji genomu opierającą się na systemie CRISPR/Cas – jeszcze nie ogólną metodę

typu „znajdź-zamień”, ale na pewno skuteczną drogę do wprowadzania w genomie dobrze wycelowanych „nacięć”. Zastosowania tej metody posypały się w niewiarygodnym tempie, nawet jak na świat biotechnologii. Już w 2013 roku dokonano udanej edycji genetycznej komórek mysich i ludzkich. W 2016 roku rozpoczęto pierwsze badania kliniczne (co oznacza, że był to konkretny, dobrze zdefiniowany i wstępnie potwierdzony program terapeutyczny, a nie „luźny eksperyment”!) nad immunoterapią genetyczną – procedurą, w toku której pobiera się komórki układu odpornościowego pacjenta, edytuje się ich genom tak, aby nauczyły się rozpoznawać wybrane zagrożenie (typowe zastosowanie to nowotwory), a następnie wpuszcza z powrotem do organizmu. A jesienią 2018 roku urodziły się Lulu i Nana – pierwsze „dzieci modyfikowane genetycznie”; ich historię opowiadam kilka akapitów niżej.

Kiedy tylko stało się jasne, że metodą CRISPR/Cas da się edytować ludzki genom, samoistnie wyłoniło się pytanie: co z embrionami? Przypuśćmy, że para ludzi ma problemy z zajściem w ciążę i korzysta ze sztucznego zapłodnienia *in vitro*. Jest taki moment, w którym lekarze mają przed sobą zygotę, z której rozwinię się niedługo człowiek. Cóż by właściwie szkodziło, żeby „dla świętego spokoju” sprawdzić jej genom, zidentyfikować w nim „proste” choroby genetyczne – na przykład opierające się na jednej nieszczęśliwej „literówce” – a następnie zrobić szybką korektę? Oczywiście, coś by też szkodziło, gdyby przy okazji udało się wyeliminować rodzinną skłonność do alkoholizmu albo tycia, albo – skoro już o tym mówimy – w ogóle przejechać przez cały genom z nieco bardziej otwartym umysłem...

No właśnie. „Cóż szkodzi”? O ile nikt nigdy nie wątpił, że dozwolona powinna być edycja genomu już narodzonych ludzi, dokonywana z myślą o likwidacji skutków chorób genetycznych, idea wyprzedzania problemu i likwidowania ich „w zarodku” od samego początku budziła sprzeciw. Po raz pierwszy udało się to zrobić na początku 2015 roku grupie chińskich naukowców^[128], którzy dokonali skutecznej edycji genomu zygoty ludzkiej (nieprzeznaczonej do implantacji kobiecie). Ich raport wywołał gigantyczną burzę w środowisku naukowym. Grupa naukowców pracujących przy inżynierii genetycznej opublikowała w „Nature” list otwarty^[129], w którym określiła edycję genetyczną komórek zarodkowych człowieka jako „niebezpieczną i etycznie nieodpowiedzialną”.

W grudniu tego samego roku zorganizowano też pierwszy Międzynarodowy Kongres na temat Edycji Ludzkiego Genomu (International Summit on Human Gene Editing), na którym przyjęto „wytyczne” dla naukowców świata, między innymi zakazujące

edytowania genów w komórkach embrionów ludzkich. Są to jednak wyłącznie „wytyczne”, a sytuacja prawna jest diablo skomplikowana. Każdy kraj ustala swoje zasady odnośnie do dozwolonych i niedozwolonych badań naukowych. Prowadzi to do sytuacji, w której w jednych państwach obowiązuje kategoryczny zakaz modyfikacji genetycznej embrionów ludzkich i za przeprowadzanie takich badań grożą poważne sankcje (jest tak w wielu krajach Unii Europejskiej, ale też na przykład w Kanadzie i Brazylii), a w innych przepisy są bardziej znuansowane – zwłaszcza w „świętej trójcy” krajów szczególnie sprzyjających takim badaniom: Stanach Zjednoczonych, Wielkiej Brytanii i Chinach. Nawet bezwarunkowy zakaz można jednak przecież złamać.

Tu dochodzimy do jedynej jak dotychczas potwierdzonego przypadku narodzenia się dzieci modyfikowanych genetycznie. 25 listopada 2018 roku chiński naukowiec He Jiankui^[85*] ogłosił, że kilka tygodni wcześniej urodziły się Lulu i Nana – dwie dziewczynki, którym na etapie zapłodnienia *in vitro* wprowadzono modyfikację w genie CCR5. Gen ten koduje białko biorące udział we wnikaniu wirusa HIV do komórek. Po modyfikacji – która uczyniła ten gen niefunkcyjnym – dziewczynki miałyby więc być rzekomo „odporne na AIDS”. 28 listopada He wystąpił na konferencji naukowej w Hongkongu, gdzie opisał swój eksperyment nieco bardziej szczegółowo (choć i tak bardzo mętnie, biorąc pod uwagę standardy akademickie). Jak łatwo się domyślić, spotkał się z niemal jednogłośnym potępieniem światowej społeczności naukowej, która żywiołowo zareagowała na jego wystąpienie już w skali kilku godzin. Sesja pytań i odpowiedzi na hongkońskiej konferencji była ostatnim publicznym wystąpieniem He.

Choć tego typu eksperymenty były wówczas zakazane w Chinach, prawo nie przewidywało określonej kary za złamanie tego zakazu. Ostatecznie He został skazany na 3 lata pozbawienia wolności i de facto wyrzucony z chińskiej i światowej społeczności akademickiej. Dziewczynki – które znamy wyłącznie z pseudonimów i przybliżonej daty urodzenia – najwyraźniej jednak żyją i mają się dobrze^[86*].

Po pierwszym szoku wywołanym przez eksperyment He od czasu do czasu zaczęły pojawiać się głosy nieco bardziej przychylnie idei „naprawiania” genomu ludzkiego na etapie zarodkowym. Bądź co bądź – można by argumentować – prowadzi się już operacje chirurgiczne na płodach ludzkich. Ponieważ zaś edycja komórek embrionalnych nie wszędzie jest zakazana – jeśli tylko nie mają one postaci pełnego embrionu, który wszczepia się kobiecie – stale

prowadzi się badania nad coraz lepszymi, coraz bardziej bezpiecznymi wersjami pierwotnej metody CRISPR/Cas.



Masowe usuwanie wad genetycznych [\[130\]](#)

Każda technologia – czy mówimy o telefonie komórkowym, czy inżynierii genetycznej – przechodzi przez z grubsze te same fazy. Najpierw jest niewygodna i niedoskonała, a dopiero później, gdy coraz więcej osób z niej korzysta, ulega uproszczeniu i udoskonaleniu. Wtedy może się upowszechnić.

W 2019 roku zespół badawczy pod kierownictwem Andrew Anzalone opisał znaczący postęp w edycji genów. Gdy opisywałem wyżej zasadniczą ideę CRISPR/Cas, wspomniałem o „podmienianiu” istniejącej sekwencji w DNA na nową, „lepszą”. Po pierwszej fali entuzjazmu to właśnie ten krok zaczął sprawiać największe problemy. Nasze komórkowe mechanizmy naprawcze są aż *zbyt* skuteczne i bywa, że zdążą zaleczyć pęknięcie w DNA jeszcze zanim zostanie w to miejsce podsunęta „lepszą” sekwencja przywleczonej razem z modułem CRISPR/Cas. Tu wkraczają autorzy cytowanego artykułu, którzy pożenili ze sobą dwie wspaniałe technologie komórkowe: przecinanie DNA (za co odpowiada CRISPR/Cas) i tworzenie nowego DNA na miejscu według wzoru (za co odpowiada białko o technicznej nazwie „odwrotna transkryptaza”).

Opisany przez nich system jest też znacznie bardziej „dyskretny”, ponieważ przewiduje jednoczesne nacięcie tylko jednej nici i to na powstałym w ten sposób „luźnym końcu” nadpisywana jest nowa informacja genetyczna – to nie budzi zaś tak wielkiego alarmu we wnętrzu komórki, co dwa luźne końce DNA. Dopiero gdy cały wzór zostanie przepisany, wykonywane jest drugie nacięcie, sprawiające, że oryginalny segment DNA zostaje usunięty. Gdyby przetłumaczyć to na naszą metaforę książki kucharskiej, metoda ta funkcjonowałaby następująco (wyobraźmy sobie ponadto, że nasza „książka” nie składa się z kartek, tylko ma postać długiej taśmy z tekstem):

1. Dostzegamy, że w przepisie na ciasto w zdaniu, które powinno brzmieć „Pieczenie mazurka rozpoczynamy od przygotowania 300 g mąki. Przygotuj też 150 g masła (lub margaryny)” zamiast „300g mąki” napisane jest „300 g kukru”. Tragedia, przecież to się skończy jakimś paskudztwem!
2. Konstruujemy nożyczki (to nasze białko Cas sprzężone z sekwencją CRISPR), które wykonują nacięcie w naszej taśmie tuż po słowach „Pieczenie mazurka rozpoczynamy od przygotowania 300 g”, oraz drugie, które wykonują takie samo nacięcie tuż przed słowami „Przygotuj też 150 g masła (lub margaryny)”.
3. Konstruujemy małą maszynkę drukarską (to odwrotna transkryptaza), która drukuje tekst na podsunięty cieniutki pasek papieru na podstawie zadanego jej wzoru. Do jej podajnika na wzory wsuwamy karteczkę z napisem „mąki”.
4. Sklejamy nożyczki z maszyną i wrzucamy do pomieszczenia z książkami kucharskimi (to jądro komórkowe).

5. Pierwsze nożyczki znajdują odpowiedni fragment i wykonują nacięcie tuż przed słowem „cukru”.
6. Maszynka drukarska zaczyna doklejać w miejscu nacięcia pasek papieru ze słowem „mąki”.
7. Drugie nożyczki wykonują nacięcie tuż za słowem „cukru”.
8. W tym momencie naturalne mechanizmy naprawcze naszej biblioteczki sprawiają, że pasek papieru ze słowem „mąki” zostaje sklejony z kartką, a swobodna karteczka z napisem „cukru” zostaje zniszczona.

Autorzy nie poprzestali na zademonstrowaniu swojej nowej technologii – która nosi, nawiasem mówiąc, nazwę *prime editing* – i postanowili pochwalić się jej medycznymi zastosowaniami. Pamiętacie pewnie, że w rozdziale tym kilkakrotnie już odwoływałem się do „prostych błędów genetycznych” o znanym podłożu i konsekwencjach? Najpopularniejszą bazą danych tego typu par „mutacja–skutek medyczny” jest ClinVar. Anzalone i współpracownicy przeanalizowali 75 122 dostępne tam wówczas „warianty genetyczne” i doszli do wniosku, że ich metoda jest w stanie „naprawić” 89% z nich. Na zachętę wykonali łącznie 175 edycji ludzkich komórek, likwidując w nich rozmaite choroby genetyczne, a także umieszczając w wybranych punktach genomu dodatkowe „dopiski” ilustrujące możliwości nowej metody. Skuteczność, zależnie od stopnia skomplikowania edycji, wynosiła mniej więcej 10–60% (czyli po poddaniu populacji komórek całej procedurze takż procent tych komórek został prawidłowo edytowany). To dobry wynik^[87]!

Dziennikarze bardzo zwięźle podsumowali to dokonanie, odwołując się do metafory komputerowego edytora tekstu. To przemawia do wyobraźni! Możemy teraz ujrzeć oczami wyobraźni, że siadamy przy komputerze przy pełnym zapisie genetycznym naszego dziecka (które jest póki co zygotą zanurzoną w jakiejś mroźnej kąpieli), po czym wspólnie z lekarzem (a może „doradcą genetycznym”, albo wręcz „projektantem genetycznym”?) dokonujemy wymarzonych zmian. Gdybyśmy projektowali film hollywoodzki, to w naszej wizji na ekranie komputera po jednej stronie znajdowałby się pewnie ciąg literek, a po drugiej wizualizacja ciała naszego dziecka w wieku 18 lat. Podmieniam parę literek i – ding! – ramiona robią się trochę szersze. Klikam w innym miejscu, a płat czołowy nieco bardziej pomarszczony. Można też do sprawy podejść jeszcze bardziej futurystycznie: przesuwam sobie różne suwaki (siła ramion, sprawność wątroby, inteligencja, skłonność do uzależnień), a równolegle w genomie wprowadzane są na bieżąco odpowiednie zmiany. Pozostaje tylko kliknąć „OK”, a w sąsiednim pomieszczeniu syntetyzowane są odpowiednie sekwencje CRISPR i „przepisy” dla odwrotnej transkryptyazy.

Cóż, jesteśmy już naprawdę blisko sytuacji, w której wszystko to byłoby wykonalne – no, może za wyjątkiem zakresu możliwości

takiego edytora i jakości „modułu wizualizacji”. Nasze braki w wiedzy na temat biologii rozwoju są po prostu zbyt duże, aby można się było w najbliższym czasie spodziewać algorytmu, który „tłumaczyłby” pomiędzy suwakami reprezentującymi sprawność fizyczną albo inteligencję a konkretnymi edycjami genomu. Jeśli jednak interesuje nas „tylko” długa lista „okienek”, na których odhaczamy sobie kolejne choroby genetyczne, przed którymi uchronimy swoje dziecko, to możemy już spokojnie sobie to wyobrażać. Dla zwiększonego realizmu należałoby jednak obok każdego okienka wyobrazić sobie również cenę. Bo nie od ograniczeń technologicznych, a prawnych i etycznych, zależy, kiedy firmy pośredniczące w sztucznym zapłodnieniu zaczną oferować „pakiety genetyczne”.

IV.3. Człowiek plus

Mój pierwszy błysk fascynacji projektem „ulepszania człowieka” (*human enhancement*) nastąpił w 2012 roku, kiedy na letnich igrzyskach olimpijskich wystąpił Oscar Pistorius. Pistoriusowi w 11 miesiącu życia amputowano obie stopy. W dorosłym życiu został sportowcem, występując z coraz lepszymi rezultatami na rozmaitych konkursach dla niepełnosprawnych. Jego protezy były jednak tak dobre, że dołączył do bardzo wąskiego grona sportowców, których dopuszczono zarówno do igrzysk paraolimpijskich, jak i „zwykłych”^[88*]. Później, niestety, okrył się niesławą: w 2013 zastrzelił swoją partnerkę, za co obecnie (lipiec 2021) odsiadyuje wieloletni wyrok więzienia.

Historia osobista Pistoriusa odciągnęła jednak uwagę opinii publicznej od ciekawego zagadnienia technologiczno-prawnosportowego: jego protezy stóp były tak dobre, że na pewnym etapie igrzysk miał realną szansę na mistrzostwo olimpijskie – nie paraolimpijskie – w biegu na 400 metrów. Co ciekawe, w igrzyskach paraolimpijskich w biegu na 200 metrów przegrał z Alanem Oliveirą, po czym w przypływie emocji oskarżył go o posiadanie „zbyt dobrych” protez – sam zresztą zmagając się już wcześniej z takim samym oskarżeniem. Ostatecznie Międzynarodowy Komitet Paraolimpijski (IPC) orzekł, że protezy Oliveiry są „proporcjonalne” i spełniają wszystkie szczegółowe wymagania techniczne. Orzeczenie to uwidocznilo jednak zasadniczy „problem” z protezami dla osób niepełnosprawnych: powoli stają się one „zbyt dobre”. Ba, nie trzeba wielkiej wyobraźni, by zauważyć, że gdyby nie istniały *żadne* ograniczenia na jakość protez, to, biorąc pod uwagę możliwości

dzisiejszej inżynierii materiałowej, w niedługim czasie zdrowi sportowcy mogliby uznać, że jedyną drogą do mistrzostwa świata jest... amputacja.

Szkoda nie wspomnieć w tym momencie o drugiej fascynującej historii, ilustrującej ideę „protezy lepszej niż oryginał”. Hugh Herr był w latach 70. obiecującym młodym wspinaczem, zdobywającym szczyty i trasy wspinaczkowe o niezwyklej trudności. W 1982 roku podczas jednej ze wspinaczek uwiązał w trudnym terenie i musiał spędzić trzy noce w śniegu, w temperaturach sięgających -30°C . Jego głęboko odmrożone nogi trzeba było amputować poniżej kolan. Wtedy zaczął się interesować protezami, ostatecznie zostając biofizykiem na Uniwersytecie Harwarda. Po kilku latach uzyskiwał we wspinaczce wyniki lepsze niż przed wypadkiem, a dziś jest jednym ze światowej klasy specjalistów w dziedzinie zaawansowanej protetyki; brał między innymi udział w tworzeniu „stóp” Pistoriusa. W 2014 roku Herr pokazał publicznie neuroprotezy, czyli protezy korzystające z naturalnych sygnałów układu nerwowego – będzie o nich mowa w rozdziale V – generujące wspaniale płynny, naturalny ruch^[131]. W 2018 roku brał już udział w serii wykładów TED z wystąpieniem pt. *Jak wszyscy staniemy się cyborgami i rozszerzymy ludzki potencjał*^[132].

Znaczna część przyjemności z pisania tej książki płynie z faktu, że każdy poruszany przeze mnie temat naturalnie prowadzi mnie ku coraz to nowszym, coraz wspanialszym nowinkom technologicznym. Ot, inżynieria sprzętu sportowego. Okazuje się, że współczesne buty do biegania, rakiety tenisowe, kostiumy pływackie i kije golfowe to prawdziwe cuda inżynierii, nierzadko produkowane staranniej niż elementy raket kosmicznych. Kryje się tu też mnóstwo problemów społecznych, jak choćby odwieczne pytanie: jak szybko biegaliby mistrzowie sprzed 50 lat, gdyby dać im wówczas dzisiejsze buty? Gdybym jednak podążał za każdym takim wątkiem, książka ta – i tak już wielotorowa – straciłaby wszelką spójność. Musicie mi więc wybaczyć, że podzielę się z wami tylko kilkoma przykładami na to, w jaki sposób inżynieria biomedyczna jest dziś w stanie zaoferować „lepszy zamiennik”.

Przyzwyczajaliśmy się już chyba trochę do „kości” z tytanu i „stawów” z polimerów wzmacnianych włóknem węglowym. Te względnie proste urządzenia mechaniczne trafiły pod strzechy i łatwo nam zaakceptować, że otrzymaliśmy w prezencie kość, która jest znacznie bardziej wytrzymała od oryginalnej^[89*]. A co powiecie na sztuczny mięsień?



Mięśnie z nanorurek^[133]

W styczniu 2021 roku opisano jeden z pierwszych funkcjonalnych prototypów mięśnia wykonanego z jednego z najbardziej wytrzymałych materiałów znanych nauce: włókna z nanorurek węglowych (*carbon nanotube yarn*). Choć niektórzy edukatorzy mówią tak jeszcze z przyzwyczajenia, od paru dekad nie jest już prawdą, że najwytrzymalsze materiały na świecie są pochodzenia naturalnego. Udało się stworzyć zarówno minerał twardszy od diamentu, jak i włókno wytrzymalsze od pajęczej nici. Zupełnie wariackie rekordy wytrzymałości pobijają zaś kolejne wersje włókien z nanorurek węglowych.

Pojedyncza nanorurka to dowolnie długi, pusty w środku cylinder zbudowany z atomów węgla ułożonych w sześciokąty. Lokalnie atomy mają więc strukturę plastra miodu, jak słynny grafen; nanorurka to więc jak gdyby długi arkusz grafenu zwinięty w tutkę. Już jedna taka rurka jest bardzo wytrzymała – przynajmniej w skali atomowej. Miliardy takich rurek, odpowiednio ze sobą splecionych (stąd „włókno”), tworzą zaś materiał o niezmiernych wręcz parametrach.

O jakiej wytrzymałości tak naprawdę mowa? Dla szybkiego porównania: w inżynierii wytrzymałość materiału na rozciąganie wyraża się poprzez siłę potrzebną do jego zerwania, przeliczoną na jednostkę powierzchni. Możemy więc wyobrazić sobie, że mocujemy do sufitu próbki wielu materiałów o takim samym przekroju i podwieszamy do nich coraz większe ciężary. W praktyce stosuje się rozmaite „rozcigarki”. Ostatecznie, ponieważ interesuje nas „krytyczna” (maksymalna) *siła*, jaką jest w stanie wytrzymać materiał o ustalonej *powierzchni*, na poziomie czysto fizycznym wychodzą z tego niutony na metr kwadratowy, czyli swojskie paskale (Pa). Mięsień ludzki ma wytrzymałość^[90*1] niecałego megapaskala (Mpa) – czyli niecały milion niutonów na metr kwadratowy. Oznacza to, że mięsień o przekroju poprzecznym 1 cm² powinien zerwać się dopiero po podwieszeniu do niego ciężarka o masie około 10 kg.

Guma ma wytrzymałość 20 Mpa. Ściegno – 80 Mpa. Kość – 130 Mpa. Nylon – 900 Mpa. Nic pajęcza 1500 Mpa, czyli 1,5 Gpa (wkraczamy w gigapaskale): włókno o przekroju poprzecznym 1 cm² utkane z nici pajęczej mogłoby teoretycznie utrzymać ciężar o masie dochodzącej do około 15 ton. Aramid, z którego robi się m.in. kamizelki kuloodporne, ma wytrzymałość 3,7 Gpa.

No i teraz włókna nanowęglowe. Ich właściwości uzależnione są od sposobu, w jaki poszczególne nanorurki zostaną splecione w jedno włókno, co jest przedmiotem żywych badań i bezcennej, pilnie strzeżonej tajemnicy handlowej. W 2019 roku ukazał się artykuł^[134], w którym podsumowano rozmaite techniki splatania opisane w literaturze i starannie zmierzono wytrzymałość otrzymanych w ten sposób włókien. Przedział uzyskanych wyników to 20–66 Gpa. Autorzy, biorąc pod uwagę dostrzeżone tendencje, szacują, że maksymalna możliwa do osiągnięcia wytrzymałość to 100–200 Gpa. Włókno nanowęglowe o tej wytrzymałości, mając przekrój poprzeczny

centymetra kwadratowego, powinno więc utrzymać ponad 1000 ton, czyli, lekko licząc, tysiąc samochodów osobowych.

Wróćmy może jednak do cytowanych badań. Oprócz niewiarygodnej wręcz wytrzymałości, nanorurka węglowa ma wielką zaletę: jest zbudowana z materiału organicznego. Nawiązuję tu do wąskiego, chemicznego znaczenia tego słowa: jako obiekty zbudowane z atomów węgla, nanorurki świetnie nadają się do tworzenia materiałów biomedycznych, „dogadujących się” z tkankami ludzkimi i innymi polimerami organicznymi stosowanymi w biotechnologii. Autorom udało się „splęść” nanorurki węglowe z włóknkami dwóch innych polimerów, które reagują na obecność jonów poprzez zmianę swojego kształtu. Wywołało to pewien spadek wytrzymałości, ale pod tym akurat względem nie brakuje marginesu. Dużo bardziej istotna jest uzyskana przy tej okazji reaktywność. Jeden z tych polimerów wywoływał po aktywacji kurczenie się całego włókna, a drugi jego rozkurczanie się. Po przyłożeniu prądu elektrycznego o niewielkim napięciu (1 V) do powstałej w ten sposób „tkanki” doszło w niej do transportu jonów, wskutek czego zaczęła się naprzemiennie kurczyć i rozciągać. Czyli... mięsień. Udało się wykonać nawet 10 tysięcy cykli skurczu bez strat jakości.

Warto podkreślić korzyści ze stosowania materiałów „zblizonych do naturalnych”: nasączony jonami polimer, reagujący na napięcie elektryczne, to coś doskonale „znanego” organizmowi ludzkiemu. Jest to wyraźna ogólniejsza tendencja w biotechnologii i inżynierii tkankowej ostatnich dekad: odchodzenie od „kółek i zębatek” w kierunku materiałów coraz bardziej podobnych do naturalnych.

Oczywiście, nie ma też nic złego w podążaniu za tendencją zgoła przeciwną. W rozdziale IV.1.2 wspominałem choćby o nanocząstkach srebra i złota. Jeżeli materiały „nienaturalne” spełniają swoją funkcję, to czemu ich nie stosować? Co więcej, nanotechnolodzy stają się ostatnio mistrzami w łączeniu ze sobą w skali atomowej materiałów „naturalnych” i „sztucznych”. To dopiero pobudzająca wyobraźnię „cyborgizacja”. Ot, choćby spójrzcie na to badanie z czerwca 2020 roku. Stanowi ono chyba najbardziej spektakularne połączenie technologii spośród wszystkich opisanych w tej książce. Pamiętam, że w trakcie lektury tego artykułu naprawdę rozdziawiła mi się paszcza – a ja od lat czytam cotygodniowo dziesiątki artykułów z bieżącej prasy naukowej. Trudno chyba o lepszy przykład na to, w jak zaawansowanych technologicznie czasach żyjemy. Magia, no po prostu magia.



Widzenie w podczerwieni [\[135\]](#)

Zadanie jest proste: przywrócić wzrok niewidomym. Przyczyn ślepoty może być wiele: zawiesić może każdy element na drodze od źrenicy, przez gałkę oczną, siatkówkę, nerw wzrokowy, aż po ośrodek wzrokowy w mózgu. W każdym przypadku interwencja musi być inna – nie ma więc jednego „leku na ślepotę”. Autorzy cytowanego tu artykułu zainteresowali się osobami, u których ślepotą spowodowana była nieprawidłowym funkcjonowaniem receptorów na siatkówce oka. W zdrowym oku w błonie komórek siatkówki „zanurzone” są cząsteczki reagujące na światło – można o nich pomyśleć jako o miniaturowych „antenkach” łapiących określony rodzaj światła. Gdy któraś z nich zostanie pobudzona, wysyła sygnał chemiczny, który później zostaje „przetłumaczony” na sygnał elektryczny wędrujący dalej nerwem wzrokowym do mózgu. Jeśli receptor taki nie funkcjonuje prawidłowo, sygnał ten nie zostaje w ogóle zainicjowany.

Autorzy opisywanego tu eksperymentu postanowili stworzyć własną „antenkę” z nanopręcików (*nanorods*) złota: to pałeczki o średnicy około 20 nm i długości około 100 nm. Co ciekawe, reagują one na światło, a zależnie od długości takiej pałeczki będzie to światło o różnej długości fali, a więc i „kolorze”. W artykule opisano nanopręciki reagujące na światło o długości fali około 900–1000 nm, czyli mieszczące się poza zakresem światła widzialnego. Inaczej mówiąc, pałeczki te pobudzają się po oświetleniu ich światłem podczerwonym.

Nanopręciki te należało następnie „zakorzeni” w błonie komórek siatkówki. Do ich powierzchni przyczepiono po kilkanaście „łapek” o ściśle określonym kształcie (czysto fizycznie, były to cząsteczki białek o nazwie immunoglobuliny, zwane też przeciwciałami) i przygotowano białka błonowe z odpowiednimi „chwytaczami łapek”. Czasem stosuje się też w takich sytuacjach metaforę zamka i klucza: gdyby ją przyjąć, trzeba by powiedzieć, że do nanopręcików doczepiono na sztorc kluczyki, a w błonie zanurzono zamki. Białka błonowe postanowiono zaś „umieścić” w siatkówce metodą genetyczną: po co żmudnie „wysiewać” je w każdej komórce albo w ogóle próbować dotrzeć fizycznie do siatkówki oka, skoro można dokonać czegoś „prostsze”: edycji genomu wszystkich tych komórek naraz tak, aby odpowiednie białko było tam po prostu produkowane lokalnie, przez same te komórki.

Do testów wykorzystano myszy posiadające mutację genetyczną całkowicie pozabawiającą je funkcjonalnych fotoreceptorów – myszy takie są, jak się łatwo domyślić, zupełnie ślepe: siatkówka ich oka nie reaguje w ogóle na jakiegokolwiek światło. Mysiom wprowadzono, posługując się specjalnie w tym celu przygotowanymi wirusami (są to tak zwane wektory wirusowe), odpowiednią modyfikację genetyczną, aby zaczęły produkować w swych siatkówkach białka błonowe, reagujące na przeciwciała sklejone z nanopręcikami. Same nanopręciki wstrzyknięto w gałkę oczną w następnym kroku. Po wykonaniu całej tej procedury myszy zaczęły prawidłowo widzieć – w podczerwieni! Przetestowano też siatkówkę oka ludzkiego – która chętnie przyjęła nanopręciki i reagowała prawidłowo na ich obecność – jednak nie przeprowadzono eksperymentów na ludziach „na żywo”: obok oczywistych

problemów etycznych warto wspomnieć, że nie przewidziano procedury „wyłączenia” tego mechanizmu widzenia, gdyby coś poszło nie tak.

Udało się też wykonać zestaw dwóch nanopręcików reagujących na różne długości fali i „zasiać” je seryjnie tak, aby biorca otrzymał odpowiednik widzenia barwnego, jednak w podczerwieni. Zwykle widzenie barwne opiera się wszak na trzech różnych typach receptorów, reagujących na różne rodzaje światła widzialnego. Gdyby stworzyć trzy rodzaje receptorów reagujących na różne długości fali podczerwonej, to mózg – któremu jest w pewnym sensie obojętne, jakiego typu receptory odpowiadają za sygnał, który trafia do niego nerwem wzrokowym, mógłby interpretować ów nowy sygnał po prostu jak różne kolory.

Przytoczone dotychczas w tym rozdziale przykłady wystarczą chyba jako demonstracja ogólnej zasady: co da się odtworzyć, da się też ulepszyć. Myślę, że pierwszymi beneficjentami tego typu „ulepszeń” będą ofiary urazów i chorób: dużo prościej jest uzasadnić etycznie i formalnie interwencję w oko czy mięsień u człowieka chorego niż zdrowego – podobnie jak trudno znaleźć lekarza, który wykona amputację zdrowej nogi sportowcowi mającemu chrapkę na supernowoczesną protezę^[91*].

Ktoś mógłby zapytać: moment, moment, a co z mózgiem? W całym tym rozdziale opowiadałem wiele o wszelakich tkankach i organach ciała ludzkiego, jednak panowała podejrzana cisza w temacie tego organu, który cały ten tekst wygenerował.

Cóż, organ ów zasługuje po prostu na osobny rozdział.

Twój umysł

W poprzednim rozdziale mówiliśmy o różnych sposobach, jakimi technologia może wnikać w nasze ciała. Wyhodowana *in vitro* nerka, rój nanorobotów medycznych w naszych żyłach, a nawet mięśnie z włókna węglowego – to wspaniałe perspektywy, bliższe niż można by przypuszczać. Choć mogą budzić pewne niepokoje, wciąż nie przekraczają pewnej subtelnej granicy – wydaje się, że scalanie się technologii z mózgiem to jednak proces w jakimś sensie odmienny od jej wnikania w organy wewnętrzne mieszczące się poniżej szyi. Sądząc po samych tytułach rozdziałów, chciałoby się powiedzieć, że mowa o różnicy między ciałem i umysłem. Sprawa jest jednak nieco bardziej subtelna, a ja chciałbym się z nią zmierzyć. Wybaczcie mi więc, że rozdział ten rozpocznie się od czegoś w stylu krótkiego eseju filozoficznego.

Z medycznego punktu widzenia niniejszy rozdział poświęcony będzie wnikaniu technologii w ludzki układ nerwowy. Rozpocniemy od interfejsów mózg-komputer, służących głównie przywracaniu sprawności osobom sparaliżowanym. W drugiej części pomówimy o scalaniu się urządzeń elektronicznych z mózgiem w celach, powiedzmy, poznawczych^[92*]: gdzieś na horyzoncie będzie nam majaczyła idea „czytania w myślach” przez elektroniczne urządzenia rejestrujące. Można by więc pomyśleć, że rozdział V tej książki mógłby być po prostu częścią rozdziału IV, poświęconą kolejnemu typowi tkanki. Czysto formalnie, to w sumie prawda, jednak nie ujmuje ona dobrze istoty problemu. Dużo lepsza, choć mniej precyzyjna, jest jednak ta intuicja filozoficzna, że wcześniej mówiłem o wnikaniu technologii w ciało, a teraz chcę pomówić o jej wnikaniu w umysł. To grząski grunt, spróbujmy jednak wyrazić tę intuicję możliwie jak najbardziej precyzyjnie.

Posiadacz organu wyhodowanego *in vitro*, a nawet właściciel genów, które zostały zaprojektowane w laboratorium jeszcze przed jego narodzinami, nie musi w ogóle o tym wiedzieć. Na co dzień żyje

w dokładnie w takim samym świecie, w jakim ludzie „w stu procentach naturalni”. Inny sposób powiedzenia tego samego: człowiek udający się na operację z udziałem technologii hodowli tkankowej, nanomedycyny czy inżynierii genetycznej po wybudzeniu się z narkozy będzie wciąż „tą samą osobą”. Jego pamięć, emocje, percepcja i sposób myślenia nie ulegną zmianie. Dopiero pod koniec poprzedniego rozdziału pojawił się pierwszy przypadek delikatnego przekroczenia tej granicy: osoba, której oczy zostałyby wyposażone w możliwość widzenia świata podczerwonego (zob. ramkę na s. 220), obudziłaby się w nieco innym „świecie przeżywanym” i prawdopodobnie operacja taka mogłaby długofalowo i na wiele subtelnych sposobów wpłynąć na jej procesy poznawcze.

W tym rozdziale poszukamy w krajobrazie współczesnej technologii takich interwencji, które mogą prowadzić do zmiany sposobu funkcjonowania umysłu, czyli, inaczej mówiąc, wpływać na nasze procesy poznawcze.

W najogólniejszym sensie *każda* technologia wpływa oczywiście na ludzki umysł. W ostatnich dekadach antropolodzy coraz odważniej i precyzyjniej mówią o koewolucji rozwoju technologicznego i neuroanatomicznego, badając na przykład, w jaki sposób na ewolucję *Homo sapiens* wpłynęło opanowanie ognia, spożywanie gotowanych warzyw i mięs, wytapianie metali i ich precyzyjna obróbka, przeniesienie się do miast, „wynalezienie” alfabetu fonetycznego albo rozwój ekonomii opartej na pieniądzu. Trudno nie wspomnieć o *Galaktyce Gutenberga* – wciąż niezwykle inspirującej książce Marshalla MacLuhana z 1962 roku, w której opisane są skutki wynalezienia alfabetu oraz druku przemysłowego dla sposobu, w jaki patrzymy na świat i o nim myślimy. Ba, na naszych oczach niemal na pewno dochodzi do kolejnej tego typu rewolucji – dokonują jej te małe, czarne ekraniki, które coraz rzadziej oddalają się od nas na więcej niż metr. W rozdziale III przedstawiłem je jako kolektory danych na nasz temat. Choć w dalszej części tego rozdziału nie będę już więcej pisał o smartfonach, warto poświęcić im choćby trochę miejsca – ponieważ ilustrują doskonale rodzaj procesów, o jakich mówię^[93*].





Twój mózg ze smartfonem

Wpływ długotrwałego stosowania smartfona na sposób funkcjonowania naszego umysłu to już niemalże wiedza powszechna – nawet jeśli nie myślimy i nie mówimy o tym na co dzień. Używanie smartfona wpływa na koordynację ręka–oko i wzorce ruszania oczami^[136], na sposób spacerowania po mieście^[137], a także na mechanizmy uwagi^[138]. Osoby, które spędzają przy telefonie wiele godzin dziennie – zwłaszcza oddziałując z mediami społecznościowymi – mają inne wzorce snu i inaczej „zarządzają” swoim stanem emocjonalnym^[139].

Myślę jednak, że tego typu przykłady wciąż jeszcze nie ujmują dobrze istoty problemu. Z badań na ten temat, które widziałem, najciekawsze – choć na pierwszy rzut oka trochę niepozorne – było chyba to^[140], w którym wykazano, że sama obecność smartfona na stole pochłania pewną część zasobów intelektualnych. Mało tego, świadome powstrzymanie się od sięgania po telefon, aby sprawdzić, „co nowego”, pochłania te zasoby jeszcze silniej. Uczestnicy badania, będący przeciętnymi użytkownikami smartfonów, zostali poproszeni o wykonanie testu mierzącego dwa aspekty zdolności poznawczych: pamięć krótkoterminową i tzw. „płynną” inteligencję, związaną z umiejętnością rozwiązywania problemów, korzystania z dostępnych danych i tak dalej. Podzielono ich losowo na trzy grupy: jedni zostali poproszeni o położenie swoich wyciszonych telefonów w widocznym miejscu na stole, ekranem do dołu; drudzy o schowanie ich w torbie, w tym samym pomieszczeniu; a trzeci o zostawienie telefonów w innym pomieszczeniu. Wyniki w obu skalach były wyraźnie lepsze wraz z rosnącym oddaleniem od telefonu. Przeprowadzony pod koniec badania kwestionariusz wykazał, że siła tego efektu zależała od stopnia „uzależnienia” danej osoby od smartfona (pytano, przykładowo, jak trudne byłoby spędzenie całego dnia bez telefonu).

Autorzy posługują się terminem *brain drain* – „drenowanie mózgu”. Cytat z artykułu: „Proponujemy, że sama obecność własnego smartfona może wywoływać «drenowanie mózgu» poprzez ciągłe obciążanie ograniczonych zasobów poznawczych”. Wygląda więc na to, że smartfon domaga się stałego udziału w naszych „zasobach obliczeniowych”, co można by wręcz przeliczyć na liczbę spalanych kalorii albo pochłanianego przez tkanki tlenu – jest to realne, stałe obciążenie dla mózgu. Gdyby stan taki trwał odpowiednio długo w skali ewolucyjnej, dobór naturalny doprowadziłby pewnie do wykształcenia się w mózgu „ośrodka monitorowania smartfona”.

Nawet i to jednak nie utrafia w istotę tego, co mam na myśli. Nie znalazłem żadnych badań, które by to dobrze wyrażały – pewnie są takie, tylko do nich nie dotarłem – posłużę się więc historią osobistą. Już jakiś czas temu zauważyłem, że gdy tłumaczę tekst angielski na język polski, odruchowo korzystam z otwartego zawsze na jednym

z „okienek” słownika komputerowego. Co więcej, robię to nawet wtedy, kiedy wiem, że znam to słowo i potrzebowałbym tylko nieco czasu, aby je „wyrzebać z pamięci”. Szybciej jest jednak sprawdzić na komputerze.

Nie znam się zbyt dobrze na neuroanatomii, jednak tak wyobrażam sobie zachodzące w mojej głowie procesy – bardziej w duchu serialu *Było sobie życie* niż poważnej kognitywistyki. Przypuśćmy, że w pewnym momencie pracy mojego mózgu pojawia się w nim zapotrzebowanie na słowo; przypuśćmy też, że chciałby je wypowiedzieć na głos. Jeżeli wytrzymam presję, by skorzystać z komputera, słowo to zostanie dostarczone z ośrodka pamięci. Jeśli jednak ośrodek pamięci nie dostarczy mi go wystarczająco szybko, to ośrodek decyzyjny postanowi posłużyć się palcami – te nacisną na odpowiednie klawisze klawiatury, a po mniej więcej sekundzie słowo pojawi się na ekranie komputera i trafi ostatecznie do mojej kory wzrokowej, a stamtąd do ośrodka „operacyjnego”. Zestawmy ze sobą te dwa procesy. W pierwszym przypadku, w uproszczeniu: zapotrzebowanie–pamięć–sygnał do strun głosowych. W drugim przypadku, w uproszczeniu: zapotrzebowanie–palce–internet–ekran–oczy–sygnał do strun głosowych. Z lekkim przymrużeniem oka można by powiedzieć, że w tym drugim przypadku fragment „moich” procesów myślowych zaszedł więc w internecie (i został odnotowany przez moją przeglądarkę, dodajmy).

Jasne – podobne procesy zachodzą od wieków, z udziałem książek, a nawet mózgów innych osób (gdy pytamy o coś, czego nie chce mi się przypominać: kiedy pracuję w tym samym pomieszczeniu, co moja żona, często współdzielimy w ten sposób zasoby mózgowe – czasem jest po prostu szybciej zapytać niż zmagać się z uczuciem „mam to na końcu języka”). Zjawisko jest jednak bardziej interesujące, gdy staje się nawykowe – gdy pewien aspekt pracy naszego umysłu zaczniemy odruchowo realizować z udziałem urządzeń elektronicznych. Dopiero w tym momencie zaczyna się autentyczne „wnikanie technologii w umysł”.

Przejdźmy może powoli do konkretów. Cały niniejszy rozdział poświęcony jest scalaniu się ludzkiej tkanki nerwowej z „maszyną”. To drugie słowo kryje za sobą w praktyce wiele możliwości. Może to być bioniczna ręka lub noga albo syntezytor mowy – o tego typu protezach dla ciała ludzkiego pomówimy najpierw, w rozdziale V.1. Może być to też po prostu bezpośrednio komputer, a za jego pośrednictwem internet, jakaś baza danych – albo terminal prowadzący do umysłu innego człowieka. O tego typu bezpośrednim połączeniu naszego mózgu ze światem informacji pomówimy w rozdziale V.2.

Kluczowym terminem dla tego rozdziału jest **interfejs mózg-komputer** (*brain-computer interface*, BCI), czyli nic innego jak połączenie mózgu z komputerem. Będzie to oznaczało albo „podpinanie się” wprost do mózgu (precyzyjniej: do ośrodkowego układu nerwowego), albo do nerwów (precyzyjniej: do obwodowego układu nerwowego). W obu przypadkach w grę wchodzi też dwa osobne procesy: pobieranie danych (odczytywanie, czyli technicznie: „obrazowanie” aktywności) oraz aktywowanie tkanki nerwowej: pobudzanie mózgu/nerwów.

W przypadku nerwów jest to znacznie prostsza procedura. Spójrz na swoją własną rękę od strony „wewnętrznej”, czyli nie grzbietu, lecz zgięcia łokcia. Być może widoczne gołym okiem są największe naczynia krwionośne przedramienia – równoległe do nich, przez sam środek ręki, tuż pod powierzchnią, biegnie nerw pośrodkowy. To naprawdę konkretny „kabelek”: przy nadgarstku ma grubość niemal pół centymetra! Nerwem tym bieżą sygnaly odpowiadające za większość zgrubnych ruchów dłoni towarzyszących na przykład zgięciu nadgarstka albo zaciśnięciu pięści. Gdyby do powierzchni twojej skóry, mniej więcej pośrodku przedramienia, przykleić **elektrodę**^[94*], to rejestrowałaby bez trudu wszystkie te sygnaly. Ba, taką samą elektrodą, a nawet źródłem pola elektrycznego oddalonym o parę centymetrów od skóry, można by „włamać się” do nerwu pośrodkowego, wysyłając za jego pośrednictwem sygnaly do dłoni. Pierwsze takie eksperymenty wykonywano już ponad 100 lat temu.

Podglądanie i aktywowanie mózgu jest znacznie bardziej skomplikowane. Po pierwsze, jest on schowany pod wielowarstwową osłoną, która sprawia, że wewnętrzne pogaduszki naszego mózgu podsłuchujemy jak gdyby zza ściany – i warstwy kołder. Po drugie, o ile nerw pośrodkowy komunikuje względnie proste sygnaly typu „Zginaj się!”, to w mózgu nakładają się na siebie tysiące elektrycznych szeptów – w języku, który dopiero zaczynamy poznawać – a konkretny głos, którego poszukujemy, bywa czasem emitowany w obszarze wielkości łąpka od szpilki, zakopanym głąboko w tkance nerwowej. Różnica między podsłuchiowaniem nerwów ruchowych a podsłuchiowaniem myśli jest trochę taka, jak między podsłuchiowaniem kaprała drącego się przy naszym uchu „Naprzód marsz!” a próbą usłyszenia, zza drzwi ruchliwej restauracji, o czym rozmawia zakochana para przy siódmym stoliku od lewej. Co nie znaczy oczywiście, że nie ma chętnych, by spróbować.

Aby zrozumieć obecny stan i perspektywy „włamywania się” do mózgu ludzkiego, musimy zacząć od szybkiego przeglądu technik sprzętowego komunikowania się z nim.





Obrazowanie i pobudzanie mózgu

Jeżeli chodzi o techniki obrazowania i pobudzania mózgu, zwyczajowo dzieli się je na inwazyjne i nieinwazyjne. Nazwa nieźle wyjaśnia istotę problemu. Techniki nieinwazyjne nie wymagają naruszenia powierzchni naszego ciała, a więc interwencji chirurgicznej. Mózgowi przyglądamy się „zza czaszki”, a nerwom przez skórę. W przypadku inwazyjnych technik badania mózgu najczęściej wykonuje się dziurę w czaszce i rozcina oponę mózgową^[95*], co stanowi, nawet dziś, potencjalnie niebezpieczną operację, której nie przeprowadza się frywolnie – jednak dzięki temu staje przed nami otworem kora mózgowa w całej okazałości.

Jak się łatwo domyślić, techniki inwazyjne pozwalają na znacznie lepszą rozdzielczość obrazowania – jednak za cenę konieczności „gmerania w mózgu”. Przez całą tę książkę przewija się temat kompromisów, ale w żadnej dziedzinie nie są one chyba tak wyraźne, co w przypadku interfejsów mózg-komputer (BCI). Nic to, działamy.

Są cztery główne techniki nieinwazyjnego obrazowania aktywności mózgu (jest też kilka innych, mniej popularnych metod obrazowania, których nie będę omawiał osobno).

Elektroencefalografia (EEG) to technika jednocześnie najstarsza i najpopularniejsza. Jej „ojciec chrzestny”, Hans Berger, dokonał pierwszego zapisu aktywności elektrycznej mózgu człowieka w 1924 roku. Metoda polega na przyklejeniu lub przyciśnięciu do powierzchni czaszki pewnej liczby elektrod, które rejestrują aktywność elektryczną „przeciekającą” z powierzchni mózgu, przez opony, czaszkę i skórę głowy. Każda elektroda rejestruje więc uśrednioną aktywność ze sporego obszaru mózgu. Zgodnie z ustanowionym w 1958 roku standardem, stosowanym do dziś na całym świecie w diagnostyce medycznej, na typowym „czepku EEG” znajduje się 21 elektrod^[96*]. Istnieje też opisana w 1985 roku wersja tego standardu o wyższej rozdzielczości, mająca 74 elektrody. Wszystko powyżej tego poziomu to już „EEG wysokiej rozdzielczości”, na przykład opisany w 2001 roku zestaw z 345 elektrodami. Nawet jednak w takim przypadku pojedyncza elektroda zbiera sygnał z regionu o objętości w najlepszym razie zbliżonej do centymetra sześciennego, a zwykle znacznie większego^[97*]. To dziesiątki i setki milionów (a bywa, że miliardy) neuronów. Szczęśliwie, „rozdzielczość czasowa” tej techniki jest bardzo wysoka – we współczesnych zaawansowanych systemach EEG sygnał rejestruje się nawet ponad 1000 razy na sekundę.

Funkcjonalna spektroskopia bliskiej podczerwieni (fNIRS) opiera się na świetle podczerwonym, którym „prześwietla się” mózg. Źródło światła i jego detektor

montuje się w pewnej odległości od siebie; w praktyce są to liczne źródła i liczne detektory, a „czepek” fNIRS wygląda ostatecznie na pierwszy rzut oka trochę jak ten do EEG. Światło podczerwone o odpowiednio dobranej długości fali względnie łatwo przenika przez ludzkie tkanki, jednak oddziałuje z barwnikami obecnymi w krwi. Co więcej, inaczej reaguje na krew natlenioną i nienatlenioną. Gdy więc mózg potrzebuje do wykonania jakiegoś zadania zwiększonej dawki tlenu w określonym obszarze, pozostawia to sygnał w detektorach fNIRS. Ponieważ światło podczerwone we wnętrzu czaszki odbija się, rozprasza i załamuje w nieprosty do przewidzenia sposób, walka o wzrost rozdzielczości nie jest łatwa – trzeba sporo zaawansowanych obliczeń, aby wywnioskować, jak właściwie powstało światło rejestrowane przez konkretny detektor, tj. jak przetłumaczyć sygnał z czujników na strukturę i aktywność mózgu. Obecnie fNIRS ma zbliżoną rozdzielczość przestrzenną do EEG – lub w najlepszym razie nieco lepszą – jednak znacznie gorszą rozdzielczość czasową.

Funkcjonalny rezonans magnetyczny (fMRI) i magnetoencefalografia (MEG) to techniki, które mają mniejsze znaczenie dla interfejsów mózg-komputer z tego prostego powodu, że urządzenia potrzebne do ich stosowania są zbyt duże, nieporęczne i kosztowne, aby można było marzyć o uczynieniu z nich produktów do codziennego stosowania. fMRI opiera się, w czym przypomina fNIRS, na badaniu przepływu krwi przez mózg – dokonuje się jednak na leżąco w tej „głośnej tubie” znanej każdemu, kto miał wykonywany rezonans magnetyczny. Nieco mniej znana magnetoencefalografia to sprytna technika, analogiczna z kolei do EEG: opiera się bowiem na wykrywaniu pola magnetycznego generowanego przez mózg przy okazji jego aktywności (każdy ruch ładunku prowadzi do powstawania pola magnetycznego). Urządzenie do wykonywania MEG przypomina przerośniętą stacjonarną suszarkę do włosów: nad głową pacjenta znajduje się gigantyczne urządzenie, w którym pracują nadprzewodniki chłodzone ciekłym helem. Pomieszczenie, w którym prowadzi się badanie, musi być też starannie chronione („ekranowane”) przed zewnętrznymi polami magnetycznymi.

Choć te „ciężkie” techniki mają gigantyczne znaczenie dla rozwoju neuronauk – a więc i również dla badania interfejsów mózg-komputer od strony teoretycznej – nie są przewidywane jako podstawa dla produktów BCI. Pewien potencjał ma fNIRS, ale gdy czyta się o faktycznie rozwijanych nieinwazyjnych interfejsach mózg-komputer, to w 90–95% przypadków opierają się na EEG. EEG jest tanie, przenośne, bezpieczne i powszechnie stosowane od dekad. Istnieją proste „czepki EEG” za kilkadziesiąt złotych, a osoba o bardzo krótkich włosach może po prostu wyjąć go z pudełka, włożyć na głowę i zacząć swoje pierwsze proste eksperymenty. Jak grzyby po deszczu pojawiają się też *jeszcze bardziej* dostępne wersje tej technologii. Istnieje choćby ear-EEG: „zauszną” wersja EEG, która wygląda trochę jak aparat słuchowy, a także OpenEEG: projekt „otwartego elektroencefalografu”, który każdy może sobie wykonać w domu (czyli mrugnęcie okiem w kierunku rozdziału II.1). Na początku 2021 roku ogłoszono prace nad tanią, przenośną wersją fNIRS o nazwie Kernel Flow.

Techniki nieinwazyjnego *pobudzania* mózgu opierają się na tych samych zasadach, co metody *obrazowania* – nic dziwnego, to przecież po prostu ta sama fizyka, tylko jak gdyby „puszczona od tyłu”. **Przecczaszkowa stymulacja magnetyczna (TMS)** polega na „szturchaniu” mózgu impulsami magnetycznymi, a **przecczaszkowa stymulacja prądem (tCS)** – impulsami elektrycznymi. Obydwie techniki mają swoje zastosowania w medycynie – na przykład w leczeniu chronicznego bólu albo depresji – jednak ograniczone póki co do oddziaływania na spore regiony w mózgu. Cóż, wygląda na to, że otaczająca mózg gruba warstwa ochronna całkiem nieźle spełnia swoje ewolucyjne zadanie, nawet w konfrontacji z technologią XXI wieku. Wszelkie sposoby komunikowania się z mózgiem spoza czaszki przypominają nieco podsłuchiwanie i podszeptywanie – wróćmy do tej metafory – zza grubej ściany wyłożonej dodatkowo

puchową kołdrą. Jeżeli jednak nie boimy się zejścia pod czaszkę, gdzie mieści się „kruchy dom duszy”^[98*], to otwierają się przed nami zupełnie nowe możliwości.

Elektrokortykografia (ECoG) to w pewnym sensie odmiana elektroencefalografii (EEG), ponieważ również opiera się na rejestrowaniu (-grafia) aktywności elektrycznej (elektro-). Człon „kortyko” oznacza zaś odwołanie do kory mózgowej. W pewnym sensie „zwykle” EEG też polega na mierzeniu aktywności elektrycznej kory mózgowej – to z tej zewnętrznej warstwy mózgu trafia wszak najwyraźniejszy sygnał do elektrod na powierzchni czaszki^[99*]. W praktyce jednak terminem „EEG” określa się metodę nieinwazyjną, a „ECoG” odnosi się do mierzenia aktywności kory bezpośrednio, elektrodami zetkniętymi z nią fizycznie. Najbardziej bezpośrednią metodą jest oczywiście „wkręcenie się” wprost w tkankę nerwową przy pomocy elektrody w kształcie igielki. W rzeczywistości dużo częściej stosowana jest metoda mimo wszystko „napowierzchniowa”, polegająca na przymocowaniu na korze, tuż pod oponami mózgowymi, kwadratowej lub prostokątnej siatki, na której znajduje się zestaw wielu elektrod w kształcie płaskich guziczków. Tego typu sieć rejestruje aktywność elektryczną na powierzchni wybranego rejonu kory. Bywa, że wybiera się metodę jeszcze bardziej bezpieczną i elektrody trafiają pod czaszkę, ale na powierzchnię opon.

W ogólności dylemat jest względnie prosty: „bliżej mózgu = lepsza rozdzielczość + większe zagrożenia i trudności techniczne” albo „dalej mózgu = gorsza rozdzielczość + mniejsze zagrożenia i trudności”. Każdy milimetr ma znaczenie, a żadna ścieżka nie pozostaje niezbadana. W 2019 roku opisano^[142] sprytną metodę polegającą na spiłowaniu kości czaszki tak, aby stała się miejscowo cieńsza, i zamocowaniu siatki mikroelektrod tylko w tym wybranym miejscu. W 2020 roku^[143] udało się zaś wprowadzić zestaw elektrod przez tętnicę szyjną i ostrożnie „podsunąć je” naczyniami krwionośnymi aż do opon mózgowych, gdzie zbierały sygnał z kory ruchowej sparaliżowanego pacjenta.

Tradycyjnie, elektrody miały średnicę liczoną w milimetrach, a odstępy między poszczególnymi elektrodami zbliżone były do centymetra. Na początku drugiej dekady XXI wieku narodziła się **mikroelektrokortykografia** (micro-ECoG albo μ -ECoG): elektrody zaczęły mieć wymiary liczone w mikrometrach (a więc mniejsze od milimetra), a z kolei ich odległość od siebie nawzajem zaczęła zbliżać się do milimetra.

W przypadku zwierząt laboratoryjnych nie obowiązują tak silne ograniczenia, jak w przypadku ludzi, i wykonywane są znacznie bardziej odważne eksperymenty. Dziś nie jest niczym niezwykłym, że do mózgu gryzonia albo małpy wszczepiana jest siatka zawierająca setki, a nawet tysiące mikroelektrod. Bywa też, że elektrodami „igielkowymi” – których czubki mają wtedy po kilkadziesiąt mikrometrów średnicy – celuje się wręcz w pojedyncze neurony. Rozwój elektrokortykografii wymaga postępu w dziesiątkach dziedzin medycyny i inżynierii. Przykładowo, na froncie materiałoznawczym ciągle trwa walka o mikroelektrody, które nie wywołują reakcji obronnej organizmu i nie zaburzają lokalnej aktywności mózgu, a na gruncie elektroniki o „okablowanie”, które jest w stanie wiernie przekazać na powierzchnię sygnał z każdej elektrody z osobna: co nam nawet i po milionach elektrod, jeśli nie będą im towarzyszyć miliony drucików? A może dałoby się ominąć ten problem przez jakiś sprytny mechanizm sumowania sygnałów z większej liczby elektrod w jednym kabelku i ich późniejszego dekodowania? Tutaj wkraczają informatycy i matematycy. Praca wre.

Gra jest jednak warta świeczki.

W 2017 opisano^[144] zestaw 96 elektrod, oddalonych od siebie o 0,7 mm (700 μ m), z których każda miała kształt kwadratu o boku 350 μ m i zbierała sygnał z obszaru kory

o powierzchni 0,12 mm^[2]. Powstały w ten sposób półprzezroczysty, silikonowy plasterek naklejono na mózg makaka w obszarze kory ruchowej odpowiadającym za jedną z dłoni. Ostatecznie udało się rozróżnić sygnały odpowiadające intencji poruszenia poszczególnymi palcami.

Jedną z zalet korzystania z elektrod jest możliwość ich stosowania na dwa sposoby – albo do pobierania sygnału, albo do jego generowania. Chociaż w praktyce nie jest to tak proste, jak przełączenie jednego pstryczka przez operatora, elektrody wszczepione pod czaszkę rzeczywiście stosuje się do stymulowania mózgu. **Głęboka stymulacja mózgu** (DBS) jest znana od lat jako metoda leczenia, choćby w przypadku choroby Parkinsona albo samoistnego drżenia rąk. W takich przypadkach stosuje się jednak raczej „klasyczne” elektrody zapewniające pobudzenie większego obszaru mózgu, nazywane potocznie „rozsusznikami mózgu”. Nie znalazłem w literaturze medycznej żadnego przypadku, gdy mózg pobudzano „precyzyjnie” przy pomocy zestawów mikroelektrod w celu wywołania efektu w skali mikroarchitektury mózgu – choć eksperymenty takie robi się z nerwami obwodowymi.

Teraz krótko o celach i zastosowaniach interfejsów mózg-komputer (BCI). Choć mogą się one wydawać futurystyczną ciekawostką, to dla wielu ludzi już dziś mają charakter konieczności. Zastosowania medyczne są też pierwsze ze względów etycznych.

Każdego roku mniej więcej 250–500 tysięcy osób na świecie doznaje **uszkodzenia rdzenia kręgowego**, głównie wskutek wypadków komunikacyjnych, upadków z wysokości i napaści fizycznej. Mniej więcej w połowie przypadków jest to uszkodzenie odcinka szyjnego, co z kolei w mniej więcej 1 przypadku na 5 prowadzi do paraliżu czterokończynowego, czyli tetraplegii^[145]. Są na świecie setki tysięcy osób, które nie potrafią poruszyć żadnym swoim mięśniem od szyi w dół – gdy z mózgu zostaje wysłany sygnał „lewa noga do przodu”, impuls elektryczny po prostu nie trafia do odpowiednich mięśni, tylko „utyka” w miejscu uszkodzenia. Istnieją rozmaite metody, którymi można próbować przywrócić takim osobom sprawność. W rozdziale IV.1.1 mówiliśmy choćby o komórkach macierzystych. Opisano już przypadki przywrócenia sprawności osobom po urazie kręgosłupa z wykorzystaniem komórek macierzystych wstrzykniętych w uszkodzony odcinek rdzenia – na całym świecie prowadzi się też na ten temat badania kliniczne^[146]. Metoda ta nie zawsze ma jednak szanse powodzenia, zwłaszcza gdy uszkodzenie jest rozległe, a organizm „oduczył się” już wysyłać prawidłowe sygnały do mięśni.

Stwardnienie zanikowe boczne (ALS) to choroba, w wyniku której stopniowo zamierają neurony ruchowe, czyli te komunikujące się z mięśniami. Osoby cierpiące na ALS stopniowo tracą zdolność do wykonywania jakichkolwiek ruchów, co zwykle rozpoczyna się od

kończyn, ale ostatecznie może doprowadzić do trudności z połykaniem i oddychaniem, a to może już doprowadzić do śmierci. Nawet gdy uda się takiej osobie zapewnić tlen i pokarm, ostatecznie może dojść do **syndromu zamknięcia** (*locked-in syndrome*), kiedy to pacjent nie jest w stanie w żaden sposób sygnalizować swoich myśli^[100*]. Stephen Hawking, u którego rozwijała się jedna z postaci ALS, w ostatniej fazie życia kontrolował wyłącznie jeden mięsień policzka i przez mniej więcej 10 lat za pomocą skurczy tego jedyne mięśnia „dyktował” swoje wypowiedzi publiczne i artykuły naukowe. Stwardnienie zanikowe boczne jest trzecią najczęściej występującą chorobą neurodegeneracyjną po chorobie Alzheimera i chorobie Parkinsona.

W obu przypadkach – urazu i ALS – w pełni funkcjonalna jest kora mózgowia, a zwłaszcza **kora ruchowa**: ten jej obszar, w którym przygotowany jest sygnał ruchowy, wysyłany następnie ku mięśniom. Co istotne, na korze ruchowej znajduje się coś w stylu „mapy ciała” – różne jej obszary aktywują się w zależności od tego, jaki ruch planujemy. Szczęśliwie, „mapa” ta ma bardzo wysoką rozdzielczość. W ramce powyżej wspomniałem o eksperymencie z 2017 roku, kiedy to udało się „odczytać” sygnał z regionów kory ruchowej makaka odpowiadających za konkretne palce – cóż za potencjał! Czysto teoretycznie, śledząc aktywność kory przy pomocy, przykładowo, elektrokortykografii, można więc dowiedzieć się, czy sparaliżowana osoba chce właśnie iść do przodu, napić się wody czy przewrócić kartkę papieru. Następnie sygnał ten można na różne sposoby wykorzystać.

Prawdopodobnie najbardziej naturalnym wyjściem byłoby pobudzenie mięśni samego pacjenta – w takim przypadku człowiek otrzymałby więc na nowo swoje własne ciało, wyposażone tylko w dodatkowe „okablowanie”. Robi się oczywiście takie eksperymenty, jednak dużym problemem jest szybkie zanikanie mięśni u osób sparaliżowanych – nie można się niestety spodziewać, aby BCI „postawił na nogi” człowieka jeżdżącego latami na wózku inwalidzkim lub przykutego do łóżka. Ponadto, elektrody „efektorowe” – czyli te przyłączone do nerwów w rękach czy nogach – musiałyby pracować bardzo „płynnie”, aby zapewnić naturalny i bezpieczny ruch. W praktyce zwykle okazuje się, że po prostu lepiej jest pobudzić do ruchu protezę ręki niż nią samą.

Stąd drugie wyjście: odwołanie się do zdobywcy robotyki: przedłużeniem ciała może stać się mechaniczna ręka, silnik wózka inwalidzkiego albo nawet egzoszkielet (zob. niżej). Jest to scenariusz,

który wyłania się też naturalnie, gdy walczymy o przywrócenie sprawności osobom z amputowanymi albo niewykształconymi kończynami. W takich przypadkach wielką nadzieją są **neuroprotezy** – czyli „aktywne” protezy rąk i nóg sterowane poprzez naturalną aktywności nerwów lub mózgu; terminem tym obejmuje się też implanty wzrokowe i słuchowe, połączone bezpośrednio z nerwami. Innym, nieco ostrożniejszym określeniem na protezy kończyn, reagujące na aktywność elektryczną nerwów, jest „**proteza mioelektryczna**”. Nazwa ta podkreśla, że mówimy o urządzeniu sterowanym przez *nerwy* w niepełnej kończynie, a nie na przykład przez korę mózgową. Pierwsze udane eksperymenty z wykorzystaniem aktywności mózgu do sterowania urządzeniem mechanicznym opisano w 1998 roku^[147]. Dodatkowa korzyść z tego kierunku badań jest taka, że tylko wyobrażenia ogranicza możliwości, jakie mogłaby uzyskać sparaliżowana osoba. Chociaż obecnie wciąż jesteśmy na etapie, na którym ciężko walczymy choćby o przywrócenie ułamka pierwotnej sprawności, nietrudno sobie wyobrazić przypadki analogiczne do historii Oscara Pistoriusa (zob. s. 215) – neuroproteza nie musi być przecież ograniczona do „naturalnej” funkcjonalności brakującej kończyny.

Trzecim wyjściem jest całkowite ominięcie ciała i w ogóle „materii”. Czasem chcemy się poruszyć tylko po to, aby przekazać lub pozyskać pewną informację. Pomyślmy przez chwilę, jak nieefektywne jest to, co robię w tej chwili: mój mózg wysyła sygnał do palców, które pukają w klawiaturę tylko po to, aby na ekranie przede mną pojawiły się literki. Za chwilę zaś te same palce pomogą mi w znalezieniu w internecie artykułu na temat najnowszych postępów w leczeniu ALS. Ciało i klawiatura są tu tylko zbędną materialną „masą oporową”, podczas gdy w obu przypadkach istota całego tego procesu dzieje się na linii umysł–komputer. Co ciekawe, nawet oś oczy–ekran komputera jest warstwą, z której można by potencjalnie zrezygnować, a i kora ruchowa nie musi brać udziału w całym tym procesie. Nie wybiegajmy jednak na razie zbyt daleko w przyszłość. Co najistotniejsze, dla wielu codziennych zadań – zwłaszcza komunikowania się z otoczeniem – tworzenie surogatu rąk, klawiatury i myszki jest zbędne, a celem jest stworzenie bezpośredniego połączenia między człowiekiem i komputerem. To, można by powiedzieć, interfejs... mózg-komputer w sensie ścisłym.

Przyjrzyjmy się może po kolei krokom na drodze ku realizacji wszystkich zarysowanych tu wizji, rozpoczynając od względnie

„bezpiecznych” przypadków kontrolowania ciała poprzez myśl, a stopniowo przechodząc ku idei „umysłu rozszerzonego”.

V.1. Mózg-komputer-ciało

Neuroprotezy to fascynujący temat, a z każdym rokiem przedstawiane są coraz „sprytniejsze” ich wersje, coraz lepiej zintegrowane z ludzkim ciałem. Zacznijmy od razu od czegoś odpowiednio spektakularnego – demonstracji maksimum naszych możliwości na dzień dzisiejszy.



Egzoszkielec sterowany myślą^[148]

W 2019 roku opisano pierwszy przypadek wyposażenia człowieka w sterowany myślą egzoszkielec, czyli „zewnątrzny szkielet”. Autorzy już w tytule zapowiadają, że jest to na razie tylko „proof-of-concept”, czyli wstępna „demonstracja wykonalności”, a nie w pełni funkcjonalny produkt. Wynik powala jednak na kolana. Pacjentowi – był nim 28-letni mężczyzna, sparaliżowany od szyi w dół^[101*] – wszczepiono dwa zestawy mikroelektrod, każdy składający się z 64 elektrod odległych od siebie o 4 milimetry. Ze względów bezpieczeństwa zdecydowano się na lokalizację na oponie mózgowej, a nie bezpośrednio na korze. Dwa zestawy, rozmieszczone na obu półkulach, wyposażono w nadajnik radiowy i bezprzewodowe zasilanie^[102*]; po operacji czaszka pacjenta została więc zamknięta i nie wystają z niej żadne kabelki.

Ostatecznie sygnał trafia do egzoszkieletu EMY (*Enhancing MobilitY*): jest to ważące 65 kg urządzenie, do którego pacjenta ciasno przyczepia się paskami. Ma ono cztery kończyny wyposażone łącznie w 14 stawów o różnym stopniu ruchliwości. Sygnał z kory ruchowej mężczyzny wysyłany jest co 100 ms (1/10 sekundy) do „plecaka” egzoszkieletu, gdzie jest on przetwarzany w sygnały do silników. Ostatecznie udało się uzyskać „czas reakcji” 350 ms, czyli zbliżony do refleksu zdrowego człowieka. Wiele czasu i wysiłku poświęcono na nauczenie programu sterującego prawidłowego odczytywania intencji pacjenta – tak, aby przelotna „myśl” o ruchu ręką (a tak naprawdę przelotna aktywność w korze ruchowej – która u zdrowego człowieka wcale nie musi prowadzić do rzeczywistego ruchu) nie zaburzała, przykładowo, trwającego właśnie spaceru. Wykonywano też wielotygodniowe testy z wykorzystaniem MEG i fMRI (zob. ramkę na s. 229), aby skalibrować urządzenie. Część tych eksperymentów prowadzona była w wirtualnej rzeczywistości: pacjent najpierw uczył się sterować egzoszkieletem na ekranie komputera, co udało mu się po mniej więcej 4 tygodniach prób. Później rozpoczęły się wielomiesięczne ćwiczenia, w czasie których opanowywał kolejne „stopnie swobody”, związane z ruchami ręką albo spacerowaniem.

Ostatecznie, do momentu publikacji artykułu sparaliżowany mężczyzna wykonał samodzielnie 480 kroków, pokonując łącznie dystans 145 metrów. Potrafił też sięgać po przedmioty i kręcić rękami w nadgarstkach – w stopniu niepozwalającym jeszcze chwycić delikatnie przedmiotów, ale umożliwiającym choćby naciśnięcie przycisku.

Muszę przyznać, że przy pisaniu tej książki niewiele tematów sprawiło mi tak wielką radość, co ten – przywracanie sprawności osobom sparaliżowanym. Jest wspaniale patrzeć na radość na twarzy sparaliżowanego od szyi człowieka, który, po prostu *chcąc* to zrobić, porusza egzoszkieletem i idzie przed siebie – nawet jeśli są to razie niezdarne, powolne kroczyki na zapewniającej bezpieczeństwo uwięzi. Trudno nie widzieć też w tych dziecięcych kroczkach tego, do czego mogą nas one doprowadzić, zwłaszcza gdy przyjrzymy się innym zastosowaniom BCI, ilustrującym znacznie bardziej subtelną kontrolę nad drobnymi partiami mięśni. Świętym Graalem jest oczywiście przywrócenie kontroli nad dłońmi lub ich robotycznym odpowiednikiem – dobra **proteza dłoni** to jednak bardziej wygórowane życzenie, niż mogłoby się wydawać.

Nie doceniamy tego na co dzień, ale dłoń ludzka jest naprawdę złożoną strukturą, która przy wykonywaniu nawet najprostszych funkcji dokonuje cudów biomechaniki. Coś tak prostego, jak naniecie się kawy z kubka, to w istocie problem inżynierski pod pewnymi względami trudniejszy od wysłania łazika na Marsa. Ludzka dłoń składa się z 27 kości, 27 stawów, 34 mięśni i ponad 100 ścięgien i więzadeł. Jest też tak naszpikowana narządami czucia powierzchniowego i głębokiego, że robotyczne dłonie wyposażone w surogat zmysłu dotyku miewają i po 100 czujników. Ktokolwiek obserwował dziecko uczące się „motoryki małej”, na przykład posługiwania się piórem albo widelcem, ten szybko orientuje się, jak wielkiej precyzji i delikatności wymagamy od naszych kończyn górnych.

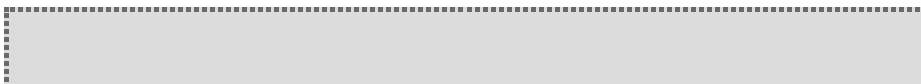
Nic więc dziwnego, że gdy tylko jest to możliwe, inżynierowie projektujący protezy dłoni starają się na różne sposoby omijać ten problem, oferując choćby urządzenia o innej budowie „anatomicznej”. Przykładowo, po co palce, skoro do chwytania świetnie nadają się, dajmy na to, przysawki? Zauważmy, że w robotyce przemysłowej nie są nigdy stosowane realistyczne, anatomicznie ludzkie palce, choć na całym świecie pracują w tej chwili miliony maszyn chwytających^[103*].

W artykule przeglądowym z 2020 roku^[150] zestawiono ze sobą 28 różnych opisanych w literaturze robotycznych dłoni z palcami. Gołym

okiem widoczne jest napięcie pomiędzy typową dla inżynierów skłonnością do upraszczania anatomii z wymogami estetycznymi. 5 palców, z czego 4 wyposażone są w po 3 paliczki każdy – to inżynierski absurd. Projektanci i użytkownicy nalegają jednak na to, aby protetyczna dłoń wyglądała naturalnie i większość protez mimo wszystko ma 5 palców; ukłonem w stronę inżynierów jest ograniczenie liczby paliczków do dwóch – to rozwiązanie jest najczęstsze. Nawet jednak tak uproszczony model stawia przed twórcami masę wyzwań – protezy te składają się z setek elementów mechanicznych, których ruchliwość musi być płynnie kontrolowana w skali milimetrów.

Wciąż istnieje niewiele protez dłoni, które pozwalałyby na naturalne i bezpieczne chwytywanie, zwłaszcza drobnych przedmiotów – o czym świadczą choćby dwa słynne Everesty robotyki: „test szklanki” i „test spinacza”. Protezy ręki obecnie radzą sobie z bezpiecznym nalaniem wody do szklanki i doprowadzeniem jej do ust oraz podniesieniem spinacza z powierzchni stołu mniej więcej tak dobrze, jak przeciętny dwulatek – z trudem i niepełną skutecznością. Z roku na rok jest jednak coraz lepiej. Firma Open Bionics specjalizuje się w lekkich, anatomicznie realistycznych (pięciopalczastych, „dwupalczkowych”) protezach mioelektrycznych rąk, drukowanych dla każdego klienta indywidualnie i dopasowanych do „stylu” jego komunikacji neuronalnej. Ich marketingowym haczykiem jest 50 wymiennych nakładek, dzięki którym proteza może naśladować rękę Iron Mana i R2-D2 ze *Star Wars* albo nawiązywać do estetyki Disneyowskiego filmu *Frozen* (najmłodszy klient tej firmy ma 8 lat). Koszt jest ustalany indywidualnie, ale podobno waha się zwykle w okolicach 10–20 tysięcy dolarów.

Poziom skomplikowania dłoni i sposobów korzystania z niej ma też jednak pewien interesujący, pozytywny skutek uboczny. Pomyślmy o pisaniu ręcznym: na przestrzeni lat nauczyliśmy nasz mózg, między innymi odpowiedni obszar kory ruchowej, „tłumaczenia” pojawiających się w naszym umyśle słów na wyrafinowany wzorzec aktywowania kilkudziesięciu mięśni. Nawet jeżeli nie posłużymy się ostatecznie fizycznie istniejącą ręką albo jej robotyczną wersją, ów obszar kory to bardzo wysokiej jakości „stacja nadawcza” naszego mózgu. Oto dwa przykłady na wykorzystanie tego potencjału.





Z myśli wprost na papier

W pierwszym badaniu^[151] testowano możliwość pisania na wyświetlającej się na ekranie komputera klawiaturze przy pomocy wyobrażonych ruchów dłonią. Uczestniczką była 61-letnia kobieta chora na stwardnienie zanikowe boczne, której wszczepiono zestaw 96 mikroelektrod w obszar kory ruchowej odpowiedzialny za ruchy prawej dłoni. Po wielomiesięcznych treningach uzyskała bardzo wysoką sprawność w przemieszczaniu kursora po ekranie komputera, ostatecznie pisząc w tempie 30 znaków na minutę. W 2017 roku, kiedy opublikowano wyniki badania, stanowiło to rekord na skalę światową. Co istotne, w badaniu tym posłużono się oprogramowaniem, które nie podpowiada słów i nie poprawia literówek – podane tempo dotyczyło przypadku „dyktowania” każdej litery z osobna. Wiadomo zaś skądinąd, że tempo pisania można znacznie podnieść, jeśli skorzysta się z algorytmów uzupełniających rozpoczęte słowa.

W drugim badaniu^[152], opublikowanym w maju 2021 roku, wziął udział sparaliżowany od szyi w dół mężczyzna. Zestaw elektrod umieszczono w jego korze ruchowej, w z grubsza tym samym obszarze, co u kobiety omówionej przed chwilą^[104*]. Poproszono go, aby wyobrażał sobie, że trzyma w ręku pióro i pisze nim po papierze. Tu wkroczyło głębokie uczenie maszynowe: pacjent na przestrzeni kilku sesji treningowych łącznie „przepisał” 572 zdania, zawierające łącznie 31 472 znaki. W miarę tego, jak na ekranie wyświetlały się kolejne zdania, mężczyzna „próbował” zapisać je ręcznie. Choć jego sparaliżowana ręka nie reagowała, kora ruchowa – owszem. Algorytm na tej podstawie nauczył się „tłumaczyć” pomiędzy aktywnością kory a literami. Ostatecznie udało się doprowadzić do sytuacji, w której pacjent „pisał” z prędkością 60–90 znaków na minutę (autorzy twierdzą, że w momencie publikowania artykułu rekord wynosił 40 znaków na minutę, został więc nieznacznie poprawiony od momentu publikacji opisywanego wyżej artykułu).

Komunikowanie się pisemne z prędkością 90 znaków na minutę, wyłącznie w oparciu o „siłę woli”, ma dla osób zagrożonych syndromem zamknięcia wprost niewyobrażalną wartość. To jednak dopiero początek tego, co jest technicznie wykonalne. Nie zapominajmy bowiem, że w mózgu inicjowane są nie tylko ruchy ręki, ale również krtani, języka, podniebienia i strun głosowych...



Generowanie mowy na podstawie intencji^[153]

Ten eksperyment wstrząsnął światem BCI w kwietniu 2019 roku. O ile stopniowe podnoszenie tempa pisania na „klawiaturze” obserwowano od lat, generowanie mowy wprost z aktywności mózgu to prawdziwa rewolucja.

Tym razem nie udało się ograniczyć wyłącznie do kory ruchowej. Mowa jest na tyle szczególną formą aktywności mózgu, że autorzy musieli rejestrować ją aż w trzech miejscach: w odpowiedniej części kory ruchowej, ale ponadto w zakręcie skroniowym górnym (tuż nad uchem) i zakręcie czołowym dolnym (nieco bliżej przodu głowy). Aby nie wszczepiać trzech osobnych zestawów elektrod, do udziału w eksperymencie zaproszono pięć osób, które już posiadały podobne zestawy elektrod w związku z leczeniem napadów padaczkowych (nie były to osoby sparaliżowane). Zestawy te nie posiadały tak wysokiej rozdzielczości, jak niektóre opisywane w tym rozdziale, jednak obejmowały wszystkie kluczowe regiony mózgu.

Większość wysiłku badaczy miało charakter obliczeniowo-informatyczny. Przyjęto strategię opartą na dwóch równie trudnych krokach:

- 1) „przetłumaczenie” aktywności mózgu na ruch aparatu mowy;
- 2) „przetłumaczenie” ruchów aparatu mowy na dźwięki.

Najpierw stworzono komputerowy model głowy ludzkiej obejmujący zwłaszcza położenie warg, języka i szczęki. Następnie uczestników poproszono, aby odczytali na głos standardowy zestaw zdań w języku angielskim, tak zaprojektowany, aby znalazły się w nim wszystkie głoski i ich typowe następstwa. Na tej podstawie wytrenowano zestaw algorytmów opierających się na głębokim uczeniu maszynowym (zob. ramka na s. 24), realizujących najpierw pierwszą część zadania – a więc odtwarzających rzeczywisty ruch aparatu mowy na podstawie aktywności kory – a następnie drugą, czyli przewidujących, jaki dźwięk produkowany jest przy danym ustawieniu anatomicznym.

W fazie zasadniczej eksperymentu uczestnicy „odczytali” 101 zupełnie nowych zdań po cichu, czyli „odczytując” słowa w myślach, czemu towarzyszą drobne ruchy aparatu mowy, tzw. subwokalizacje. Algorytm generował na tej podstawie nagranie dźwiękowe. Następnie przeprowadzono badanie online z udziałem ochotników, którym prezentowano pojedyncze słowa wraz z listą 25 lub 50 słów do wyboru. Wyniki były różne: dla pewnych zdań osiągnęły 100% (wszyscy słuchacze prawidłowo zrozumieli wszystkie słowa), dla innych tylko 30%. Warto podkreślić, że w zestawie pojawiły się również zdania naprawdę trudne, zawierające rzadkie słowa i zbitki głosek (przykład z badania: „Etiquette mandates compliance with existing regulations”^[105*]). Co więcej, przyjęta metoda oznaczała, że słuchacze byli całkowicie pozbawieni kontekstu – autorzy postawili więc poprzeczkę naprawdę wysoko.

Posłuchałem nagrań z tego eksperymentu: jakość głosu była zbliżona do tego, jaki wydają na przykład osoby leczące się po wylewie krwi do mózgu albo zmagające się z porażeniem mózgowym lub poważnym przypadkiem jąkania. Przeciągnięte samogłoski, niewyraźne spółgłoski, głos dość „bełkotliwy”. Ktokolwiek jednak miał do czynienia z osobami mówiącymi w ten sposób, wie też, że do owego specyficznego „lejącego się” głosu można się dość szybko przyzwyczaić. Nie zapominajmy też, że

w rzeczywistych sytuacjach komunikacyjnych w zrozumieniu drugiego człowieka bardzo pomaga kontekst, my zaś póki co walczymy o to, aby całkowicie sparaliżowana osoba mogła powiedzieć choćby: „Cześć!” albo „Co tam nowego u Stefana?”. Nie bądźmy zachłanni – w końcu, podkreślmy to, mówimy o urządzeniu, które *mówi za sparaliżowaną osobę, czytając jej w myślach!*

Rozdział ten póki co wypada chyba szalenie poważnie. Cóż, zaczęliśmy od perspektywy medycznej, ponieważ rozwój BCI i neuroprotezy jest w dużym stopniu napędzany przez chęć przywrócenia sprawności. Nie brakuje jednak wizjonerów, którzy myślą już o rozszerzaniu możliwości osób najzupełniej sprawnych. W poprzednim rozdziale wspominałem o próbach ulepszenia ciała ludzkiego – przez superszybkie protezy nóg, nanocząstki zapewniające widzenie w podczerwieni albo mięśnie z włókna nanowęglowego. A czy dałoby się analogicznie ulepszyć nasz umysł? Póki co temat ten jest głównie przedmiotem prowokacji artystycznych – nie zaszkodzi jednak przyrzec im się, choćby pobieżnie.



Cyborg art

Hiszpański artysta Manel Muñoz, znany pod pseudonimem Manel de Aguas, zainstalował sobie nad uszami dwie „płetwy”, zawierające w środku czujniki temperatury, wilgotności i ciśnienia atmosferycznego. Sygnał jest następnie tłumaczony na wzorec drgań implantu wszczepionego do jego kości skroniowej. Drgania kości czaszki są rejestrowane przez narządy słuchu – to tzw. słyszenie przez przewodnictwo kostne – tak więc można powiedzieć, że Muñoz „słyszy stan atmosfery”. Choć czysto technicznie nie jest to BCI, Muñoz należy do najbardziej znanych przedstawicieli *cyborg art*: ruchu artystyczno-technologicznego, którego zasadniczą ideą jest rozszerzenie ludzkiej perspektywy zmysłowej w celach poznawczych i artystycznych.

Katalońska tancerka Moon Ribas – znajoma i współpracowniczka Muñoz – wszczepiła sobie z kolei w stopy generator wibracji, który łączy się bezprzewodowo z globalną siecią sejsmometrów, informując ją drganiami o trwających właśnie trzęsieniach ziemi. *Waiting for Earthquakes* to kompozycja taneczna Ribas, która rozpoczyna się od stania w bezruchu w oczekiwaniu na trzęsienie ziemi^[106*]. Choreografia zależy już od przebiegu drgań w trakcie konkretnego epizodu sejsmicznego – kobieta tańczy więc wraz z wibrującym globem, na którym stoi.

„Ojcem założycielem” ruchu *cyborg art* jest jednak Neil Harbisson, który od 2004 roku posiada na stałe wszczepioną w czaszkę antenę rejestrującą kolory, a następnie tłumaczącą je na wzorzec drgań czaszki – zupełnie jak u Muñoza. Harbisson jest więc „technologicznym synestetykiem” – terminem „synestezja” określa się bowiem występujące u niektórych ludzi „przepływanie” pomiędzy zmysłami, na przykład „smakowanie kolorów” (widzeniu kolorów towarzyszą wrażenia smakowe) albo „widzenie dźwięku” (dźwiękom towarzyszą wrażenia barwne). Po latach spędzonych z dodatkowym zmysłem mężczyzna ten żyje już w świecie, w którym dźwięk kostnego brzęczenia i kolory są na trwałe zespolone w jego świadomości. Istnieją niezliczone kreatywne zastosowania tego urządzenia: Harbisson komponuje na przykład „muzyczne portrety”, inspirując się dźwiękiem generowanym przez jego urządzenie, gdy czujnik przesuwany jest przed twarzą danej osoby.

Wibrator Harbissona połączony jest dodatkowo z odbiornikiem radiowym i może zdalnie przyjmować sygnały dźwiękowe (i barwne). Jest to więc nie tylko „dźwiękowy czujnik barw”, ale ponadto bardziej uniwersalne urządzenie odbiorcze. Mężczyzna jest na przykład w stanie odbierać połączenia telefoniczne, których dźwięk nadawany jest poprzez przewodnictwo kostne. Przez pewien czas Harbisson pozwolił też grupie przyjaciół wysyłać mu zdalnie „pocztówki kolorowe”. W 2014 roku opowiadał, że gdy wiadomość taka przychodzi w środku nocy, sny otrzymują odpowiednie zabarwienie.

Harbisson i Ribas – którzy są z kolei przyjaciółmi z dzieciństwa – zamontowali też sobie Bluetooth Tooth: implanty zębowe pozwalające im na dyskretną formę komunikacji. Każde z nich ma jeden taki ząb, zawierający maleńki wibrator, odbiornik/nadajnik Bluetooth oraz niewielki przycisk. Delikatne przygryzienie tego przycisku powoduje przesłanie drgnięcia do zęba drugiej osoby, co pozwala im na komunikowanie się alfabetem Morse’a.

Choć tego typu projekty artystyczne mogą wydawać się ledwie zabawą, zwłaszcza w zestawieniu z wytężonymi pracami nad coraz to lepszymi neuroprotezami, to jednak ich nie bagatelizujemy. Przedstawiciele „cyborg art” poruszają bowiem szalenie ważny temat: prawdziwe scalenie umysłu i technologii ma nastąpić dopiero wtedy, kiedy urządzenia elektroniczne połączone z naszą tkanką nerwową staną się czymś więcej niż tylko zdalnie sterowanym narzędziem – kiedy staną się „częścią nas”. Harbisson tak zżył się już ze swoją anteną – stale sterczącą nad jego głową jak owadzi czulek – że traktuje ją jako część swojego ciała. W 2004 stoczył batalię z brytyjskim urzędem paszportowym, aby ten zaakceptował obecność anteny na jego zdjęciu paszportowym. Wymogi formalnie zakazują, aby na fotografii znajdowały się urządzenia elektroniczne, jednak Harbisson tak długo przekonywał urzędników, że jest to część jego ciała, że ostatecznie dali mu spokój. W 2011 roku jego antena została uszkodzona w trakcie demonstracji w Barcelonie przez policjanta, który myślał, że jest to urządzenie rejestrujące obraz. Harbisson, konsekwentnie, oskarżył go o uszkodzenie ciała – a nie zniszczenie własności.

Wbrew pozorom, za tego typu prowokacjami artystycznymi kryje się ważna myśl: odpowiednio silnie scalone urządzenie elektroniczne rzeczywiście może stać się „częścią mnie”. Antena Harbissona jest już jednym z jego narządów zmysłów. A co, gdyby – kto wie, być może pomysł ten został już przez kogoś zrealizowany? – zamontować na niej dodatkowo czujniki rejestrujące uszkodzenia fizyczne, przekazujące sygnał do ośrodków bólu? W takim przypadku szarpiący się z anteną policjant wywoływałby też fizyczny ból. Wspomniany w rozdziale IV.3 Hugh Herr w trakcie jednego ze swoich wykładów zaproponował proste kryterium pozwalające odróżnić, kiedy „człowiek z protezą” staje się „**cyborgiem**”. Następuje to wtedy, gdy proteza nie tylko odbiera sygnały i reaguje na nie, ale ponadto zwraca sygnały do układu nerwowego właściciela, efektywnie stając się częścią jego „odczuwanego ciała”. Co istotne, są dobre powody, by rozwijać ten kierunek badań: detektory nacisku w protezie ręki, przykładowo, dostarczają bezcennych informacji o jakości chwytu i właściwościach dotykanych przedmiotów. Póki co informacje te są zwykle przetwarzane „lokalnie” i służą lepszej kontroli nad palcami – dlaczego jednak nie mielibyśmy dołączyć do neuroprotezy dodatkowego zestawu elektrod, podłączonych do nerwów *wychodzących* z kończyny? Są już pierwsze eksperymenty tego typu.



Proteza, którą czujesz

W kwietniu 2020 roku opisano^[154] wyniki kilkuletniego eksperymentu z udziałem czterech ochotników z rękami amputowanymi powyżej łokcia. Było to pierwsze doniesienie o udanym wieloletnim posługiwaniu się w warunkach życia codziennego protezą, która komunikuje się z układem nerwowym „dwukierunkowo”. Z jednej strony była to zwykła pięciopalcza proteza mioelektryczna, reagująca na inicjowane w mózgu sygnały, wędrujące nerwami ruchowymi do amputowanej kończyny. Wyposażono ją jednak dodatkowo w trzy czujniki nacisku, zamontowane po wewnętrznej stronie protezy dłoni. Sygnał z nich trafiał do nerwów czuciowych, przenoszących go ostatecznie do mózgu. Czujniki te stały się więc w pewnym sensie narządami zmysłu pacjentów, informującymi ich o nacisku na „ich” dłonie.

Tego typu eksperymenty muszą trwać latami. Zupełnie jak w przypadku „artystów cyborgów”, układ nerwowy musi „nauczyć się” obustronnej komunikacji i do niej przyzwycząić, aby urządzenie mechaniczne naprawdę zaczęło funkcjonować i być

postrzegane przez mózg jak część ciała. Wygląda jednak na to, że jest to możliwe. Uczestnicy eksperymentu zgłaszali, że stopniowo wraca im zmysł dotyku, początkowo opisując go jako „dotknięcie czubkiem długopisu” zlokalizowane w protetycznej dłoni. Z czasem udało się sprawić, że zmienny stopień nacisku w trzech czujnikach był odczuwany przez pacjentów jako „elektryczne drganie” o różnej intensywności, co pozwoliło na wykonywanie coraz bardziej wyrafinowanych czynności.

Jest istotne, abyśmy wyraźnie podkreślili, co tak naprawdę zaszło w toku tego eksperymentu. Zostały bowiem przekroczone granice tego, co określa się jako „moje ciało”.

Zauważmy, że gdy moja dłoń trzyma jakiś przedmiot, nie muszę na nią zerkać – jak na imadło albo szczyptce – aby wiedzieć, że uchwyt jest solidny: po prostu *wiem to, czuję to*, ponieważ dłoń jest częścią mojego ciała. Uczestnicy opisywanego wyżej badania również czuli, że „ich dłoń” wykonała solidny chwyt, a gdy wyczuwali, że przedmiot się wysuwa, mogli zacieśnić uchwyt, wysyłając do neuroprotezy odpowiedni sygnał nerwami ruchowymi. Czysto technicznie, jednokierunkowa komunikacja (układ nerwowy→proteza) została „po prostu” zastąpiona komunikacją dwukierunkową (układ nerwowy↔proteza), co może z pozoru brzmieć jak drobiazg technologiczny. Według definicji Herra ta czwórka osób stała się jednak cyborgami.

Wspomniałem wyżej o tym, że „cyborgowym artystom” nie potrzeba receptorów bólu; poza oczywiście potrzebą wyłącznie artystyczną. Dla właścicieli dwustronnych neuroprotez mogłyby one tymczasem pełnić funkcję praktyczną. Zauważmy, jak istotne są one choćby w trakcie gotowania; umiejętność wyczuwania uszkodzenia „tkanek” prowadziłaby też do dłuższej żywotności protez. Dopóki nie są one produkowane z materiałów, które nasz organizm traktuje jak własne i naprawia na bieżąco, dobrze byłoby poczuć, że „plastik mojego przedramienia” zaczyna się trochę przypalać – ból zapewniałby więc protezie większą trwałość. W końcu bardziej dbalibyśmy o coś, czego uszkodzenie fizycznie nas boli.

Nie natrafiłem jeszcze na ten pomysł w literaturze naukowej, ale można by się zastanowić, czy receptory bólu w protezach nie powinny aby generować własnego „łuku nerwowego”, analogicznego do bezwarunkowego odruchu cofnięcia ręki (może pamiętacie jeszcze z rozdziału I.2.1, że programowanie działań „oddechowych” jako osobnego, szybszego mechanizmu reagowania, nie jest niczym niezwykłym). W takim przypadku bioniczna ręka sama odskakiwałaby

po jej zetknięciu z rozgrzanym brzegiem patelni, a użytkownik pozbawiony byłby przykrych skutków ubocznych w postaci doświadczenia bólu. Sama korzyść praktyczna z bólu bez tego, co w nim nieprzyjemne. Jedyne problemy polegałyby na tym, że osoba wyposażona w taką protezę od dzieciństwa nie nauczyłaby się ostrożności – bioniczna ręka odskakiwałaby, ale mózg nie kojarzyłby nieostrożnego zachowania z czymś nieprzyjemnym. Rozważania możemy pociągnąć oczywiście dalej: gdyby proteza ta była wyposażona w swoją sieć neuronową uczącą się na błędach (jak roboty w ramce na s. 28), być może wykształciłby się w niej jakiś lokalny surogat ostrożności, niepołączony z naszym mózgiem. Czy jednak chcielibyśmy mu pozwolić na wykonywanie prewencyjnych ruchów zapobiegających oparzeniu? To już ingerencja w nasze zachowanie nieco poważniejsza niż odsunięcie się od ognia. Ech, zostawmy może tę fantazję na boku.

Przywrócenie sprawności ciała – czy nawet rozszerzenie jego funkcjonalności – to nie koniec. Zauważmy, że naturalnym środowiskiem dla komputera nie jest świat materii, lecz świat informacji. Poruszenie choćby i jednym robotycznym palcem oznacza dla komputera konieczność „wyjścia ze strefy komfortu”, podczas gdy wykonanie milionów operacji na gigantycznej bazie danych albo przeszukanie całej Biblioteki Kongresu w poszukiwaniu słowa „kalafonia” przychodzi mu naturalnie i z wdziękiem. Z drugiej strony my, ludzie, z każdym stuleciem coraz mniej polegamy w życiu codziennym na naszych ciałach, a rośnie znaczenie naszych umysłów. W dzisiejszych czasach można spędzić cały dzień, nie wykonując prawie żadnych ruchów mięśni, w międzyczasie zarabiając pieniądze, komunikując się z bliskimi i oddając się rozrywce. Zdrowe to może nie jest, ale z pewnością coraz bardziej realne dla rosnącej rzeszy Ziemiaków. Nic więc dziwnego, że wielkim zainteresowaniem cieszy się wizja scalenia człowieka z komputerem nie w celu podbicia świata materii, lecz świata informacji i wiedzy.

V.2. Mózg-komputer (bo po co ciało?)

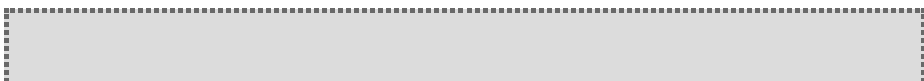
Zauważmy, że w poprzednim rozdziale zarysowała się dość ciekawa procedura „wracania do korzeni” po tajemniczych ścieżkach naszej tkanki neuronalnej. Stosunkowo najbardziej „niewinnym” wynalazkiem są neuroprotezy, które podłączamy na samym czubku naszej sieci nerwowej – ot, choćby na 40 centymetrach brakującej

kończyny. Jeszcze bliżej źródła jest rdzeń kręgowy, następnie pień mózgu, a ostatecznie sam mózg. Zależnie od typu aktywności można jednak pokusić się o bardziej szczegółową analizę. Przykładowo, gdy mam ochotę sięgnąć ręką po kubek kawy, czynność ta najpierw jest planowana w przedniej części kory czołowej, a dopiero później wyłania się „plan ruchowy” w leżącej bliżej czubka głowy korze ruchowej. Dzięki wysokiej rozdzielczości technik obrazowania takich jak fMRI (zob. ramkę na s. 229) jesteśmy już dziś w stanie zarejestrować samą intencję wykonania jakiejś czynności, nie tylko zanim drgnie czyjaś ręka, nie tylko zanim sygnał ruchowy trafi do rdzenia, ale nawet zanim ktoś *uświadomi sobie*, że podjął jakąś decyzję.

Pomyślmy teraz o sytuacji, w której poruszenie ciałem jest tak naprawdę niepotrzebne. Jeśli moim celem jest, przykładowo, sterowanie urządzeniami domowymi albo kierowanie samochodem, to moje ręce i nogi są zbędnym pośrednikiem na drodze między mózgiem a jakimś urządzeniem, które pewnie i tak byłoby szczęśliwsze, gdyby uruchamiał je sygnał elektroniczny, a nie ludzki paluch albo buciór.

W 2019 roku opisano proste urządzenie^[155] – oparte na „czepku” EEG, a więc niewymagające otwierania czaszki – które rejestrowało intencję awaryjnego hamowania samochodu. Badani siadali przed symulatorem jazdy po mieście, wyposażonym w ekran, kierownicę i zestaw pedałów. W pewnym momencie widoczny na ekranie pojazd zaczynał ostro hamować tuż przed maską. Sytuacja wymagała awaryjnego zatrzymania, co następowało średnio po 880 ms (0,88 s) od momentu zapalenia się świateł hamowania w widocznym na ekranie samochodzie do momentu, gdy pedał hamulca zaczął wykrywać nacisk. Analiza sygnału elektrycznego z mózgu pozwalała zaś na przewidywanie, że użytkownik planuje hamować, średnio 600 ms wcześniej. Mózgowi zajęło więc tak naprawdę zaledwie 280 ms sformułowanie wyraźnej „intencji zahamowania” od momentu, gdy światło hamowania dotarło do siatkówki oka. Pozostałe 600 ms to żmudne tłumaczenie tej intencji na „plan ruchowy”, a następnie wędrówka przez mózdzek i rdzeń, aż do nerwów i ostatecznie pobudzenie mięśni nogi do skurczu.

Bezpieczeństwo to oczywiście rzecz ważna, jednak nie aż tak ważna jak telepatyczne sterowanie światłem w pokoju i robotem domowym. Nareszcie coś praktycznego!





„Przynieś mi piwo, robocie”, pomyślałem

W 2017 roku opisano^[156] prosty system służący do kontroli minimalnych urządzeń domowych (żarówka i wiatraka) przy pomocy intencji, wychwyconej za pośrednictwem EEG. Autorom udało się nauczyć algorytm, aby rozpoznawał, o którym z tych dwóch przedmiotów myśli użytkownik, a za komendę „włącz/wyłącz” posłużyło, co ciekawe, mrugnięcie oczami – wywołujące na tyle charakterystyczny wzorec aktywności, że funkcjonowało jako „kliknięcie”. Projekt ten miał charakter dość prostej demonstracji i prawdopodobnie nie sprawdzilby się jeszcze w warunkach rzeczywistych, natomiast – co istotne – urządzenie wykonano z tanich, powszechnie dostępnych komponentów. Przykładowo, rolę EEG pełnił komercyjnie dostępny zestaw MindWave zbierający dane wyłącznie z fragmentu kory czołowej (kosztujący około 800 zł w momencie pisania tych słów), a rolę komputera spełnił „minikomputer” Raspberry Pi.

Nieco bardziej zaawansowany jest projekt rozwijany od lat przez grupę australijskich naukowców na Uniwersytecie Nowej Południowej Walii. W najnowszej wersji ich urządzenie^[157] ma już stanowić uniwersalną platformę do sterowania „sprytnym domem” (*smart home*) – a więc systemami oświetleniowymi i klimatyzacją, a także sprzętami takimi jak odkurzacze czy nawet roboty domowe. Podstawą sprzętową jest nieco bardziej profesjonalny „czepek” EEG Emotiv EPOC+ (do kupienia za około 4000 zł), a algorytmy oparte na głębokim uczeniu maszynowym długo trenowano z udziałem grupy ochotników. Sporo wysiłku poświęcono na stworzenie systemu dyktowania liter – o tym zadaniu jednak mówiliśmy już wyżej. Drugim obszarem zastosowania jest sterowanie robotem domowym, póki co niestety dopiero wirtualnym. Autorzy pokazują jednak, że dobrze wytrenowany algorytm jest w stanie zrozumieć człowieka, który wysłał mentalny komunikat „Podaj napój”, a także odczytać bardziej szczegółowe myśli, za pomocą których użytkownik kieruje robotem.

Im bardziej subtelna aktywność umysłowa, tym trudniej o jej wykrycie. Różnicę pomiędzy odczuwaniem gniewu, myśleniem o przyszłości oraz słuchaniem muzyki jest w stanie wykryć nawet prosty zestaw EEG – te trzy procesy umysłowe aktywują bowiem wyraźnie odmienne regiony mózgu. Docelowo chcielibyśmy jednak pewnie, aby interfejs mózg-komputer odczytywał konkretną myśl, na przykład „Ciekawe, jakich rozmiarów jest bakteria *Escherichia coli*”. W takim razie komputer mógłby wyszukać dla nas tę informację w internecie – co jest już możliwe – a następnie ją nam podpowiedzieć – najprostszym wyjściem byłoby posłużenie się

przewodnictwem kostnym, jak Neil Harbisson, dzięki czemu słyszelibyśmy odpowiedź w postaci cichutkiego głosu rozlegającego się po prostu „w głowie”. Jedynym brakującym elementem układanki, aby mógł powstać system bezpośredniego mózgowego dostępu do wszelkiej wiedzy, jest – bagatela – precyzyjnie rozpoznawanie treści myśli na podstawie analizy aktywności mózgu w czasie rzeczywistym.

Tu wracamy do wszystkich problemów i kompromisów zarysowanych w ramce na stronie 229. Z każdym rokiem jesteśmy w stanie identyfikować coraz lepiej myśli ludzkie, jednak z reguły wymaga to rozdzielczości właściwej dla fMRI. EEG nie wystarczy. Fascynujące jest śledzenie doniesień literaturowych na temat „czytania umysłu” przez komputery. Choć nie jest łatwo odczytać krążące nam po głowie myśli – i do dziś nie ma żadnego wiarygodnego doniesienia na ten temat – dużo wyraźniejsze są jednak neuronalne ślady po oglądanych przez nas, lub choćby tylko wyobrażanych sobie, obrazach.

Już w 2005 roku udało się^[158] wykryć – właśnie przy pomocy fMRI – czy badana osoba myśli o wizerunku twarzy, miejsca czy przedmiotu. Eksperyment początkowo zaprojektowano odważnie: uczestnikom pokazano 10 zdjęć twarzy słynnych osób, 10 zdjęć znanych miejsc i 10 zdjęć przedmiotów spotykanych w życiu codziennym. Następnie – już w tubie fMRI – kazano im wrócić pamięcią do konkretnej fotografii. Ostatecznie udało się wykryć tylko ogólną kategorię, do której należał wyobrażany obraz, a nie konkretnie, czy dana osoba myśli o wieży Eiffla, czy o Koloseum. To było jednak w 2005 roku. Choć sama rozdzielczość fMRI trochę się od tego czasu polepszyła, dużo bardziej spektakularne były w tym czasie postępy w rozwoju uczenia maszynowego.

W 2018 roku opublikowano wynik eksperymentu^[159], w którym wytrenowano głęboką sieć neuronową na wzorcach aktywności kory wzrokowej, rejestrowanych przez fMRI, w momencie oglądania różnego rodzaju obrazów. Powstał w ten sposób algorytm potrafiący „tłumaczyć” pomiędzy prostym czarno-białym obrazem a aktywnością kory mózgowej. Ten sam algorytm „puszczony wspak” (technicznie, posłużono się siecią typu GAN – zob. ramkę na s. 120) był w stanie przewidywać, na jaki obraz patrzy dana osoba wyłącznie na podstawie danych z fMRI. Co istotne, na tym etapie eksperymentu uczestnicy patrzyli na obrazy, których *nie widzieli* w fazie treningowej: algorytm więc nie tyle wybierał jeden spośród wielu znanych mu obrazów, co naprawdę generował czarno-biały obraz na podstawie wzorca aktywności mózgu. Jakość tych rekonstrukcji była różna – czasem

obraz zrekonstruowany odpowiadał blisko oryginałowi, czasem jednak system zaliczał spektakularną wpadkę.

Badania takie mają wielkie znaczenie dla rozwoju BCI, ponieważ kora wzrokowa może pełnić dla myśli podobną rolę, co kora ruchowa dla intencji ruchowych – jako wysokiej rozdzielczości „ujście” dla czegoś, co jest głęboko ukryte w mózgu w postaci morza subtelnych sygnałów. Ba, pewne kluczowe dla procesów myślowych obszary mózgu są dosłownie „zakopane” głęboko pod powierzchnią, jak choćby hipokamp biorący udział w procesach pamięciowych. Sygnały jego aktywności elektrycznej kompletnie się rozmywają po drodze na powierzchnię i szalenie trudno jest je analizować przy pomocy EEG. Próby umieszczania mikroelektrod w głębokich regionach mózgu, na przykład na hipokampie, wykonywane są zaś w ostateczności i tylko w celach praktycznych, na przykład w trakcie operacji^[160].

Jak pamiętamy z poprzedniego rozdziału, da się odczytać słowa, o których myślimy, jeśli tylko spróbujemy je „przetłumaczyć” na ich zapis na papierze. Teraz wiemy, że na podstawie aktywności kory wzrokowej da się odtworzyć oglądane lub wyobrażane przez kogoś zdjęcie – z każdym rokiem z coraz lepszą „rozdzielczością”. My, ludzie, jesteśmy naprawdę wzrokowymi stworzeniami, i naszemu intensywnemu myśleniu lubi towarzyszyć aktywność kory wzrokowej – a już na pewno, kiedy próbujemy sobie coś wyobrazić. Być może w przyszłości wydawanie poleceń BCI będzie wymagało nauczania się nieco bardziej wizualnego stylu myślenia?

Na koniec dwa ciekawe eksperymenty przeprowadzone przez garstkę zapaleńców, głównie z Uniwersytetu Helsińskiego: wedle mojej wiedzy jest to jedyna grupa badawcza na świecie, która używa terminu **interfejs mózg-informacja** (*brain-information interface*). Jesteśmy jeszcze daleko od odczytywania konkretnego przedmiotu naszych myśli na podstawie aktywności mózgu, jednak problem ten można próbować obejść. Kluczowe jest samo dostrzeżenie problemu.



A może to by cię zaciękawilo?^[161]

Autorzy tego badania postanowili stworzyć interfejs mózg-komputer, który pomaga w nawigowaniu po oceanie informacji dostępnych w internecie. Ich rozumowanie jest następujące: gdy wchodzimy w kontakt z jakąś treścią na ekranie komputera – na przykład artykułem na Wikipedii – część tekstu będzie dla nas bardziej interesująca, a część mniej. Próbując zgłębić jakieś zagadnienie – albo po prostu miło spędzić wieczór, czytając na ulubiony temat – będziemy poszukiwać dalszych treści tylko na tematy istotne/interesujące. W skali sekund i minut robimy to, po prostu klikając w odpowiednie linki lub poszukując dalszych informacji poprzez wyszukiwarke. Wikipedia, będąca tu poligonem doświadczalnym, jest wprost naszpikowana linkami – podążanie za głodem wiedzy jest więc technicznie proste, choć można też utonąć w zalewie możliwości. Drugi problem pojawia się w dłuższej skali czasowej: czyż nie byłoby świetnie, gdyby nasz komputer już z góry wiedział, co będzie dla nas istotne?

Tu wkraczają fińscy badacze. Uczestnicy ich eksperymentu siadali przed komputerem z czepkiem EEG na głowie. Z listy 30 tematów (bankowość, Michael Jackson, komunizm, koty, schizofrenia, wino...) mogli wybrać dowolny, który uznali za interesujący. Na ekranie pojawiał się odpowiedni artykuł z angielskiej Wikipedii, słowo po słowie^[107*], a aktywność ich mózgu analizowana była ze względu na to, jak silny był sygnał „zainteresowania-istotności” po ujrzeniu kolejnego słowa^[108*]. Program odczytujący poziom zainteresowania danej osoby kalibrowano wcześniej, indywidualnie dla każdego uczestnika, poprzez przedstawienie obok siebie zdań z artykułów uznanych za daną osobę za interesujące i nieinteresujące.

Po przeczytaniu przez użytkownika pewnej partii tekstu algorytm miał do dyspozycji listę „obiecujących wątków” – słów, które wydawały się szczególnie pobudzać użytkownika. Z niej wybierano te, które prowadziły do dalszych artykułów na Wikipedii – w ten sposób powstał system „rekomendacji na żywo”. Autorzy docelowo przewidują, że system taki mógłby służyć jako asystent w procesie zbierania informacji, odciążając nas z „ciężkiej pracy” klikania na linki i korzystania z wyszukiwarki.

Drugim zastosowaniem takiego systemu jest długoterminowe uczenie się zainteresowań użytkownika. Autorzy lamentują, że wszelkie istniejące systemy rekomendacji – filmów, muzyki, artykułów – opierają się albo na naszych własnych ocenach typu „Lubię to!” (z czego użytkownicy, jak pokazują badania, korzystają bardzo niekonsekwentnie) albo na ich rzeczywistej aktywności w internecie. Choć to drugie źródło informacji jest bezcenne (o czym mówiliśmy sporo w rozdziale III.3), pozostaje mimo wszystko bardzo „płaskie” w porównaniu z żywą reakcją naszego mózgu. Klikamy na setki linków, jednak nie wszystko, do czego prowadzą, równo nas interesuje. Pewnej informacji dostarcza czas naszego oddziaływania z daną treścią – który jest oczywiście pilnie odnotowywany, analizowany i gromadzony – jednak nic nie powie tak szczerze, co nas *naprawdę* interesuje, jak nasz własny mózg. Hm, gdy się nad tym chwilę zastanowić, to stan naszego mózgu jest nie tyle *sygnałem* naszego poziomu zainteresowania, co po prostu *jest naszym zainteresowaniem*.

Autorzy powyższego artykułu nie przedstawiają żadnej konkretnej wizji *produktu* powstałego na bazie ich systemu, nietrudno jednak wyobrazić sobie program komputerowy, który, przykładowo, na górze ekranu wyświetla czytany aktualnie dokument, a w dolnej części

sugestie pozyskane na podstawie sygnału z EEG. W duchu „łączenia kropek” trudno nie wspomnieć też o omówionych w rozdziale II.2.1 generatorach tekstu, które mogłyby przecież otrzymywać na bieżąco *feedback* na temat poziomu naszego zainteresowania, generując dla nas opowieść albo agregując informacje ze źródeł naukowych. Tego typu generator, „podpięty” pod czepek EEG, mógłby sugerować nam na próbę różne wątki, wyczuwając, czy jesteśmy nimi zainteresowani, a następnie spełniając nasze niewypowiedziane na głos oczekiwania. Trochę w ten sposób – znów pobawmy się kropkami – działają wszak algorytmy podsuwające nam reklamy na podstawie zgromadzonych na nasz temat „wielkich danych” (o czym mówiliśmy w rozdziale III.3).

A teraz jeszcze sprytniejsze wykorzystanie tego samego sprzętu – no i generatywnych sieci adversarialnych (zob. ramkę na s. 120). Jak widać, nasi dzielni fińscy badacze nie zasypiają gruszek w popiele. Takich pionierów nam trzeba!



Tworzenie myśla^[162]

Oto i prawdziwe połączenie kropek. We wrześniu 2020 roku helsińscy naukowcy opisali program komputerowy, który generuje na żywo obraz kierowany wyłącznie sygnałem z EEG.

Podstawą jest opisany wyżej algorytm przewidujący poziom istotności/zainteresowania. Tym razem uczestnikom eksperymentu postawiono zadanie polegające na wygenerowaniu myślami określonego rodzaju twarzy: na przykład uśmiechniętej, albo z blond włosami, albo młodej, albo kobiecej. Na ekranie wyświetlały się przypadkowo wygenerowane twarze – łącznie 70, co pół sekundy. Jedynym zadaniem uczestników było wpatrywanie się w nie i przypominanie sobie od czasu do czasu o tym, jaka ma być „ich” twarz. Okazuje się, że ich mózgi reagowały inaczej na zdjęcia „pasujące” ze względu na określone kryterium niż na zdjęcia „niepasujące” – a więc gdy, przykładowo, zadanie polegało na myśleniu o twarzy uśmiechniętej, to EEG rejestrowało specyficzny sygnał w czasie oglądania zdjęć ludzi uśmiechniętych: bez względu na ich wiek, płeć, kolor skóry, włosów i tak dalej.

W tle tymczasem „pracował” sobie algorytm generatywny, bezustannie wprowadzając drobne korekty do jednego tylko obrazu ludzkiej twarzy. Był to dokładnie taki sam algorytm typu GAN, o jakim była mowa w rozdziale III.3 – zdolny więc do generowania fotorealistycznych wyników. Tym razem algorytm ten dostawał jednak na bieżąco informacje, czy docelowa twarz ma być bardziej czy mniej podobna do każdego kolejnego obrazu oglądanego właśnie przez uczestnika

eksperymentu. To wystarczyło i po 35 sekundach takiego eksperymentu sieć neuronowa produkowała twarz z pożądaną cechą w 90% przypadków.

I... chyba tyle. Muszę zawieść tych z was, którzy spodziewali się, że w tym rozdziale opiszę interfejsy mózg-komputer czytające w naszych myślach albo „dopisujące” do naszego mózgu nową wiedzę. Tego typu urządzeń jeszcze nie ma, choć niektórzy nadmiernie entuzjastyczni dziennikarze od czasu do czasu sugerują, że coś takiego właśnie powstało^[163].

Spokojnie, jeszcze nie. Brakuje jeszcze paru kropek.

Na koniec

Jeżeli po przeczytaniu pięciu poprzednich rozdziałów kręci wam się trochę w głowie i zadajecie sobie pytanie, czy z tego wszystkiego wyłania się w ogóle jakikolwiek spójny obraz przyszłości – to dobrze. Przyszłość nie będzie spójna, podobnie jak spójna nie jest teraźniejszość.

Pamiętajcie – nie szukamy w tej książce „wielkiej narracji”, podporządkowanej rozwojowi jednej technologii, uporządkowanej i sensownej. Pobawmy się trochę przyszłością i skorzystajmy z tego poczucia wolności, jaki odczuwa się po zanurzeniu w chaosie. Zresztą, przecież naprawdę w przyszłości wydarzyć się mogą rzeczy najprzedziwniejsze!

Na kartach tej książki próbowałem was od czasu do czasu namówić na zabawę w łączenie kropek, skacząc między rozdziałami i łącząc ze sobą na próbę technologie, które rzadko omawia się ramię w ramię. Jeżeli 59 znajdujących się w tej książce ramek potraktować jako kropki na mapie przyszłości, to można je wszak połączyć na dowolny sposób – i za każdym razem wyłoni się z tego inny obraz, nawet jeśli nieco surrealistyczny. Spróbujmy może pobawić się razem w futurologiczną ruletkę!

Na zachętę przedstawię wam mój atak na trzy tego typu pary, wybrane metodą losowania.

1. Architektury kognitywne (s. 49) i automatyczna synteza chemiczna (s. 86).

Hm, co by się stało, gdyby urządzenie do syntezy chemicznej „spiąć” z jakąś zaawansowaną architekturą kognitywną? Gdyby architektura ta była ściśle racjonalna, można by ją stosować do przeprowadzania optymalizacji syntezy. To jednak nudne, potrzebujemy czegoś ciekawszego... A co, gdyby posłużyć się którąś z architektur wyposażonych w moduł nudy albo moduł poszukiwania nowych bodźców? Tego typu „mały chemik” mógłby wytwarzać nowe związki chemiczne bez żadnego zainteresowania ich praktycznym zastosowaniem, kierując się na przykład ich kolorem albo szybkością zachodzenia reakcji. A może urządzenie takie mogłoby wręcz „celowo” doprowadzić do wybuchu?

Inną możliwością byłyby „sprzęgnięcie” tych dwóch urządzeń na sposób chemiczny. Może pamiętacie, że niektóre architektury kognitywne mają „układ hormonalny” – na stronie 59 opisałem przypadek wyposażenia robota w wirtualną adrenalinę. A czemu nie mielibyśmy stworzyć architektury kognitywnej reagującej na rzeczywiste hormony i neuroprzekaźniki? Potrzebne by było urządzenie, które jest w stanie naprawdę reagować na substancje chemiczne tak samo, jak ludzka tkanka nerwowa, czyli – patrzcie, jak się dobrze składa! – organoid mózgowy na płytce (s. 207).

Wyłąnia się z tego wszystkiego następujące urządzenie: architektura kognitywna sprzęgnięta z prostym „mózdzkiem”, zbudowanym z prawdziwej tkanki nerwowej, która dzięki temu może reagować na sygnały chemiczne. Owa architektura steruje ponadto urządzeniem do automatycznej syntezy chemicznej („chemputerem”), które jest w stanie wstrzykiwać produkty reakcji do zbiornika, w którym znajduje się organoid mózgowy. Ostatecznie powstał więc „komputer chemiczny”, który jest w stanie samodzielnie generować sobie substancje psychoaktywne, a następnie przeprowadzać kolejne rundy syntezy już pod ich wpływem. Hm. Nie wiem, czy to ma sens.

2. Drony zabójcy (s. 35) i pijana kobieta podnosząca ciężary (s. 125).

Brzmi jak wyzwanie. Jedyne, co mi przychodzi do głowy, to wyposażenie robota – nie wiem, czy miałyby to sens w przypadku drona – w możliwość naśladowania ruchów ludzkich. W rozdziale I.1 pisałem trochę o staraniach, aby roboty poruszały się jak najbardziej naturalnie. Algorytm tego typu, opisany w ramce na stronie 125, mógłby pozwolić robotom jeszcze bardziej „wtopić się w tłum” – jest on wszak w stanie wygenerować sekwencję ruchów na hasło „zmęczony człowiek uginający się pod ciężarem plecaka” albo „radosna dziewczynka podskakująca w biegu”. Czyżby robot zabójca miałby do nas przyjść krokiem Clinta Eastwooda?

3. Tworzenie myśla (s. 255) i masowe usuwanie wad genetycznych (s. 212)

Też niełatwe zadanie. Być może pomysł, żeby losować te tematy, nie był najlepszy? Spróbujmy jednak.

Idea tworzenia myśla jest taka, że osoba siedząca przed ekranem reaguje spontanicznie na wyświetlające się na nim treści, podczas gdy analizowany stan jej mózgu służy komputerowi do odgadywania, „o co chodzi” tej osobie. W pewnym sensie jest to więc swego rodzaju

detektor kłamstw – również kłamstw, które sami sobie opowiadamy. Przypuśćmy, że powstaje program analogiczny do opisanego w ramce na stronie 255, którego zadaniem jest wybranie dla mnie idealnej partnerki. Na ekranie wyświetlane są setki zdjęć, a sygnał z czepka EEG służy do wykrywania, które z tych zdjęć budzą moje najsilniejsze zaciekawienie. To samo można zrobić z tembrem głosu, przykładowymi wypowiedziami, listą zainteresowań oraz ostatnio przeczytanych książek i tak dalej. Byłaby to z pewnością ciekawa nakładka na serwisy randkowe – myślę, że wiele osób mogłoby się zdziwić, pozytywnie lub nie, zobaczywszy, jaki jest dla nich idealny partner zdaniem ich własnego mózgu, nieprzefiltrowanego przez konwencje społeczne, wstyd albo przyzwyczajenia.

Czy ma to jakiegokolwiek przełożenie na edycję genomu? Pierwsza rzecz, która przychodzi mi do głowy, to zastosowanie analogicznej procedury do procesu „projektowania dziecka”, o którym wspominałem w rozdziale IV.2. Być może przyniosłoby to jakąś korzyść, gdyby przyszli rodzice przeszli przez proces zautomatyzowanego odczytania nadziei wiązanych ze swoim potomstwem? A może w głębi duszy pragną, aby ich dziecko było chuderlawym molem książkowym – i zarówno oni, jak i ich dziecko, byłiby w takim razie szczęśliwsi – ale pod presją plakatów wiszących w poczekalni przed gabinetem genetycznego projektowania dzieci zdecydowałiby się na model mięśniak numer 5?

Ech, fantazje, fantazje...

Mam nadzieję, że odnajdywanie tego typu połączeń sprawi wam choć trochę frajdy, a może nawet znajdziecie mnie gdzieś, w internecie lub osobiście, i podzielicie się ze mną swoimi wizjami przyszłości.

Spokojnie, będzie dobrze.

Spis ramek

*Numery stron odnoszą się do wydania papierowego.
Na potrzeby e-booka mają zadane linkowanie i kierują do
odpowiedniej ramki w tekście.*

I. Twój robot

Poćwicz ze mną! [18](#)
Dolina niesamowitości [21](#)
Uczenie przez wzmacnianie [24](#)
Sam się nauczę chodzić [28](#)
Rój badawczych dronów [32](#)
Drony zabójcy [35](#)
Płyn robotów [37](#)
Modularny pomocnik domowy [38](#)
Boty konwersacyjne i test Turinga [45](#)
Architektury kognitywne [49](#)
DIARC [51](#)
Dempster i Shafer [57](#)
O robocie, który się bał [59](#)
Robot ciekawski [62](#)
Widziałem to na filmie [66](#)

II. Twój świat

Druk 3D [81](#)
Automatyczna synteza chemiczna [86](#)
Rolnictwo komórkowe [91](#)
Słowa jako wektory [99](#)
Mądrość miliona tekstów [107](#)
AI Dungeon [110](#)
Project Debater, czyli jak się pokłócić z komputerem [112](#)
Konwolucyjne sieci neuronowe [117](#)
Generatywne sieci adversarialne [120](#)

Pijana kobieta podnosi ciężary [125](#)

AIVA [126](#)

III. Twoje dane

Centra danych [136](#)

Kable podmorskie [139](#)

Deep web, dark web, Tor [143](#)

Metadane [151](#)

Co Google wie [156](#)

Automatyczny test osobowości [162](#)

Facebook cię zdiagnozuje [164](#)

IV. Twoje ciało

Komórki macierzyste [173](#)

Inżynieria tkankowa [177](#)

Oskrzele z hodowli komórkowej [178](#)

Pusty organ czeka na komórki [179](#)

Mysia tarczyca z drukarki 3D [182](#)

Nanokapsułka z lekiem reagująca na 5 bodźców [187](#)

Origami DNA [189](#)

Nanorobot DNA walczy z zakrzepami [191](#)

Bakteria dostarcza lek do guza nowotworowego [197](#)

Organoidy [203](#)

Mózdzek w hodowli zaczyna się budzić [204](#)

Ciało na płytce [207](#)

CRISPR/Cas [209](#)

Masowe usuwanie wad genetycznych [212](#)

Mięśnie z nanorurek [218](#)

Widzenie w podczzerwieni [220](#)

V. Twój umysł

Twój mózg ze smartfonem [225](#)

Obrazowanie i pobudzanie mózgu [229](#)

Egzoszkielec sterowany myślą [237](#)

Z myśli wprost na papier [241](#)

Generowanie mowy na podstawie intencji [242](#)

Cyborg art [243](#)

Proteza, którą czujesz [246](#)

„Przynieś mi piwo, robocie”, pomyślałem [250](#)

A może to by cię zaciekało? [253](#)

Tworzenie myślą [255](#)

Przypisy

[1*]. Wszystkie cytowane tu dane podaję za raportem ONZ: *World Population Prospects 2019*. Twórcy tego raportu opracowali kilka symulacji przyszłego rozwoju światowej populacji; ja podaję wartości za „względnie ostrożnym” wariantem o nazwie technicznej *medium variant*.

[2*]. Ściślej: urodzeń żywych.

[3*]. Mobilność reprodukcyjna mężczyźni nie ma tu znaczenia. Dopóki kobiety stanowią około 50% populacji, dopóty każda z nich musiałaby urodzić dwójkę dzieci, aby populacja utrzymywała się na stałym poziomie.

[4*]. Słowo „algorytm” oznacza po prostu ustalony sposób wykonywania jakiegoś zadania. Algorytmem jest na przykład sekwencja: rozgrzej patelnię – wbij jajko – mieszaj, aż się zetnie. W kontekście komputerów algorytm jest programem komputerowym.

[5*]. W chwili pisania tej książki model podstawowy kosztuje 6 tysięcy dolarów, a wybór rozmaitych podnoszących jakość produktu dodatków szybko powoduje wzrost wyceny do 10 tysięcy i więcej.

[6*]. Polskie tłumaczenie tego terminu, niestety już upowszechnione, jest trochę niefortunne. Słowo *uncanny* oznacza raczej „niepokojący”, „budzący nieufność”. Oto piękne wyjaśnienie z *Oxford English Dictionary*, monumentalnego słownika języka angielskiego: „o osobach: taki, któremu nie jest bezpiecznie zaufać lub wchodzić z nim w kontakt ze względu na jego związki ze sztukami lub mocami ponadnaturalnymi; [...] tajemniczy, dziwny, nieprzyjemnie dziwny lub obcy”. Lepsza byłaby chyba jednak „dolina niepokoją”, a może wręcz „niespokojna dolina”?

[7*]. W momencie publikacji tej książki na przykład *Genetic Algorithm Walkers* (https://rednuht.org/genetic_walkers/) oraz *Evolution by Keiwan* (<https://keiwan.itch.io/evolution>) – w przypadku tego drugiego użytkownik sam decyduje o anatomii swojego robota. Obserwowanie uczących się chodzić dwunogów w tej pierwszej symulacji to pouczająca lekcja siły i słabości algorytmów genetycznych.

[8*]. Najprostsze sieci neuronowe „wypluwają” po prostu pewien zestaw sygnałów wyjścia („reakcja”) w odpowiedzi na bieżący stan sygnałów na wejściu („bodźce”). W latach 80. i 90. zaczęły nie miało powstawać, a w XXI wieku upowszechniać się, tzw. **rekurencyjne sieci neuronowe** (*recurrent neural network*), posiadające dostęp do historii swojego własnego stanu – a więc coś w stylu „pamięci”. Oczywiście, każda sieć neuronowa, która przechodzi przez proces uczenia się, ma z samego tego powodu swego rodzaju „pamięć” zakodowaną we własnej strukturze. Można by o niej pomyśleć jako o pamięci długoterminowej, albo po prostu o zbiorze posiadanych umiejętności. Jeżeli jednak mamy do czynienia z siecią już „nauczoną”, czyli taką, w której układ i właściwości neuronów nie ulegają dalszej zmianie, to w najprostszym przypadku będzie ona zawsze reagowała na ten sam zestaw wejść w ten sam sposób: będzie „ślepa na historię”. Dopiero uwzględnienie odwołania do przeszłości w samej strukturze sieci zapewnia coś w stylu „pamięci krótkoterminowej”. W pewnych zadaniach jest ona wręcz niezbędna. Pomyślmy o odczytywaniu tekstu pisanego ręcznie. Sieć neuronowa „bez pamięci” podchodzi do każdego znaku jak do osobnego zadania do rozwiązywania. Tymczasem wiedza o tym, jakie były poprzednie znaki, daje wielką moc przewidywania. Dokładnie taki sam niezbytelną gryzmoł po literach „po godzinac...” to najprawdopodobniej „h”, a po literach „trudny przypade...” to najprawdopodobniej „k”.

W ostatnich latach dodawanie do sieci neuronowych kolejnych warstw „pamięci” przyniosło znaczący postęp w rozwiązywaniu najróżniejszych problemów, w tym również i chodzenia.

[9*] To z kolei tak zwane **sieci adwersarialne** (*adversarial networks*), z których największą sławą okryły się adwersarialne sieci generatywne (GAN), które omawiam w ramce na stronie 120. Zdaję sobie sprawę, że wprowadzam tu bardzo dużo terminów technicznych, ale staram się ograniczać tylko do tych, które najprawdopodobniej będą się „pojawiać tu i tam”, gdy zainteresuje was któryś z tematów opisanych w książce.

[10*] Pod koniec XX wieku opisano tzw. model SLAM (Simultaneous Localization and Mapping, czyli „jednoczesne lokalizowanie i mapowanie”), a w pierwszej dekadzie tego wieku pojawiły się jego pierwsze przekonujące realizacje. Gdy szuka się informacji na temat autonomicznych robotów mapujących, warto poprobać z hasłem „SLAM”.

[11*] Angielskie *quadrocopter* jest coraz częściej spolszczane jako „kwadrokopter”, ale ten potworek nie przejdzie mi przez usta.

[12*] Stosują się tu zwyczajne „prawa” wyścigu zbrojeń, a rozmiary dronów i ich względnie niska cena mogą być naprawdę kłopotliwe. W marcu 2021 roku serwis „The War Zone” opisał wyniki dochodzenia w sprawie tajemniczego roju dronów, które w lipcu 2019 roku spędziły kilka dni, krążąc wokół amerykańskiego niszczyciela USS Kidd – najwyraźniej całkowicie autonomicznie. Ostatecznie, przynajmniej na poziomie publicznie dostępnych dokumentów, pochodzenie dronów nie zostało zidentyfikowane^[15].

[13*] Można sobie z nią pogadać pod adresem: <https://chat.kuki.ai/chat>.

[14*] Słowo „architektura” kojarzy się dość jednoznacznie z projektowaniem budynków, więc z początku może brzmieć dziwnie w przytoczonym tu zastosowaniu. O „architekturach kognitywnych” można by też pewnie mówić jako o „schematach kognitywnych” albo „strukturach kognitywnych”.

[15*] Autorzy pracy posługiwali się algorytmem rozpoznawania mowy Google Speech API, który nie obsługuje słów spoza słownika, jednak wyposażony jest w gigantyczny słownik uaktualniany na bieżąco przez „szperacze” internetowe – w praktyce oznacza to, że dopóki uczestnicy opisywanego tu eksperymentu nie próbowali celowo wymyślać dziwacznych słów, robot nie miał problemu z ich zrozumieniem. Oczywiście, istnieją też algorytmy, które obsługują nieznanne słowa. Gdy pojawia się nowe słowo, użytkownik może zostać na przykład poproszony o jego przeliterowanie, po czym trafia ono na stałe do słownika.

[16*] Osobom szukającym nowej muzyki, która ma w sobie choć trochę elektroniki, z całego serca polecam Ishkur's Guide to Electronic Music: <https://music.ishkur.com/>. Gdyby udało mi się kiedyś ogarnąć naukę i technologię w stopniu, w jakim Ishkur ogarnia muzykę elektroniczną, mógłbym spokojnie przejść na emeryturę i zająć się hodowlą pszczoł.

[17*] Istnieje też handlowa nazwa Fused Deposition Modeling (FDM), czyli dosłownie: „modelowanie poprzez osadzanie stopionego [materiału]”, która bywa czasem używana zamiennie z FFF. Wygląda na to, że twórcy czuli niedosyt trzyliterowych skrótów.

[18*] Istnieją, nawiasem mówiąc, „zaczarowane ołówki”, którymi przy odrobinie wprawy można „rysować w przestrzeni” plastikowe rzeźby – można o nich pomyśleć jako o drukarkach typu FFF, w których użytkownik trzyma głowicę w ręku.

[19*] Wymowa: „mejker”. To, niestety już upowszechniona, kalka z angielskiego *maker movement*.

[20*] Pierwsi kosmonauci rzeczywiście trenowali jedzenie z tubki – ponieważ w latach 60. tak właśnie sobie wyobrażano jedzenie w kosmosie. W międzyczasie okazało się, że na stacjach kosmicznych znajdzie się miejsce i dla niezbędnego sprzętu, i dla koktajli krewetkowych, i dla ekspresu do kawy. Ekspres ISSpresso, zaprojektowany specjalnie z myślą o pracy w stanie nieważkości, stworzyły dla Międzynarodowej Stacji Kosmicznej (ISS) firmy Argotec i Lavazza. Pierwsze świeżo zaparzone espresso w kosmosie wypiła 3 maja 2015 roku włoska (oczywiście...) kosmonautka Samantha Cristoforetti.

[21*] W momencie pisania tego tekstu nie ma już ich w menu – być może pandemia podcięła tej restauracji skrzydła? Są natomiast dostępne *impossible burgers*, czyli wegetariańskie substytuty mięsa z produktów roślinnych – jest to jednak zupełnie inny koncept, choć też ciekawy i mogący silnie wpłynąć na rynek gastronomiczny.

[22*] Czysto technicznie, jajo składa się z komórek w tym sensie, że samo jest początkowo jedną wielką komórką, a po zapłodnieniu powstają w nim kolejne komórki embrionu. Jajko w sensie gastronomicznym, tj. białko + żółtko, nie składa się jednak z komórek – są to raczej niekomórkowe „materiały biologiczne” wytworzone dla potrzeb embrionu.

[23*] Przez pierwsze dwa lata istnienia, czyli w latach 2014–2016, występowała pod nazwą Muufri, od *muu-free milk*, czyli „mleko wolne od muu”. Może nikt nie łapał tego żartu? Ja się ubawiłem.

[24*] Plant nigdy nie ukrywał, że jest to tylko hobby, a on sam na co dzień żyje w zwykłym domu, korzystając w pełni z dobrodziejstw cywilizacji.

[25*] Techniczne określenie na opisywaną tu procedurę to *word embedding*, czyli „osadzanie słów”. Sama idea polega jednak na „zamianie słów na wektory”, więc ja w tej książce będę mówił po prostu jako o „wektoryzacji słów”, a nie ich „osadzaniu”, co jest chyba nieco klarowniejsze.

[26*] Popularnym wyborem przy trenowaniu algorytmów jest cała angielska Wikipedia, zawierająca obecnie około 4 miliardów słów – gdyby ją wydrukować w postaci opasłych tomów wielkości Britanniki, zajęłaby ich około 3 tysięcy.

[27*] Wersja testowa NKJP dostępna jest w internecie pod adresem www.nkjp.pl. Wszystkie przykłady w tym rozdziale to rzeczywiste wyniki z NKJP.

[28*] Nazwa tego projektu została utworzona od środkowego członu słowa „encyclopedia” i jest w związku z tym wymawiana „sajk”.

[29*] Jego polską wersją jest *Va banque*.

[30*] Tekst, również w wersji ustrukturyzowanej, gdzie wyjaśnione jest, jak powstało każde zdanie, znajduje się w cytowanym wyżej artykule naukowym (Slonim i in., 2021). Pełne nagranie debaty można w momencie publikacji tej książki obejrzeć tutaj: <https://www.youtube.com/watch?v=m3u-1vttrVw>.

[31*] Gdybyście się zastanawiali, to owa procedura „nakładania” na siebie dwóch obrazów i sprawdzenia, na ile ze sobą „harmonizują”, to właśnie „konwolucja”. Czysto matematycznie, jest to operacja wykonywana na dwóch funkcjach, zwyczajowo określana w polskich podręcznikach matematycznych terminem „plot”. Gdyby więc zachować to nazewnictwo, *convolutional network* należałoby przetłumaczyć jako „sieci splotowe”. Krajowe odpowiedniki nazw dla nowych technologii powstają jednak szybko, najczęściej jako proste kalki z języka angielskiego, a kiedy puryści i eksperci zaczynają wyrabiać sobie zdanie na temat jakości tłumaczeń, nazwy zwykle są już dawno w praktycznym użytku.

[32*] Jeżeli chcemy, aby nasza sieć wyszukiwała krzesła, możemy oczywiście stworzyć serię filtrów „krzesłowych”. Można też tworzyć sieci konwolucyjne poszukujące liter, twarzy i tak

dalej. My niżej skupimy się na sieciach ogólnego przeznaczenia. Druga uwaga: ilustracje w tej książce są czarno-białe, ale sieci konwolucyjne w ogólności „pracują” na obrazach kolorowych, a rzeczywiste zestawy filtrów zawierają zwykle również filtry kolorowe, na przykład „zielona smuga nad niebieską smugą”.

[33*] Technicznie, jest tak, ponieważ filtr „przykładany” jest do obrazu wejściowego z pewnym „skokiem”. Założmy, że zaczynamy od obrazu o szerokości 1000 pikseli i badamy go ze względu na lokalne podobieństwo do filtra, który sam w sobie ma 50 pikseli szerokości. Nie przykładamy go jednak w każdym możliwym miejscu na obrazie, lecz kolejne próby oddalone są od siebie o, powiedzmy, 10 pikseli. W takim razie mapa uzyskana przy pomocy pierwszego filtra będzie miała tylko 100 pikseli szerokości (i wysokości). (Uważni czytelnicy mogą się pewnie zastanawiać, czy metodą tą nie zgubi się wzorów występujących tuż przy brzegu obrazu. Usłusznie informuję, że zwyczajowo otacza się obraz pustym „marginosem” o wymiarze połowy szerokości filtra. Tego typu subtelności jest więcej, ale żadnym sposobem nie dałoby się ich tu wszystkich opisać).

[34*] Choć współczesne sieci typu GAN nie są tak naprawdę po prostu „odwróconymi na opak sieciami konwolucyjnymi”, to metafora ta może być źródłem pożytecznych intuicji. Zauważmy, że sieć konwolucyjna może określić tysiące różnych obrazów poprzez tę samą sekwencję słów, w dodatku zrobić to z obrazem, którego nigdy wcześniej nie widziała (do tego została wszak stworzona!). Po odwróceniu kierunku przepływu informacji okaże się więc, że z tej samej sekwencji słów mogą zostać uzyskane tysiące różnych obrazów, wytworzonych „od zera”: nie na podstawie bazy danych treningowych, tylko informacji zakodowanych w ostatniej warstwie „konwolucyjnego tortu”.

[35*] Już sama konieczność istnienia tego typu detektorów fałszywek pokazuje, jak dobre one są. Inny ciekawy kierunek przeciwdziałania fałszywym treściom to tym silniejsze potwierdzanie autentyczności prawdziwych treści. W momencie wydawania tej książki firma Truepic współpracuje z gigantem telekomunikacyjnym Qualcomm nad stworzeniem „nakładki” na popularny system operacyjny Android, dzięki której wykonane smartfonem zdjęcia i filmy otrzymywałyby „certyfikat autentyczności”, przypisany do konkretnego urządzenia rejestrującego.

[36*] Linuksowym odpowiednikiem wiersza poleceń jest Terminal, a odpowiednia komenda to „traceroute” – tak zresztą brzmiała w starszych Windowsach.

[37*] Dla dodatkowego szpanu można to zrobić, wciskając jednocześnie klawisz Windows i „R”, a następnie wpisując „CMD”.

[38*] Czysto technicznie, „przechowywanie w chmurze” (*cloud data storage*) ma oczywiście swoje określone znaczenie. Idea „chmury” polega na tym, że właściciel i użytkownicy danych nie odwołują się do nich bezpośrednio (czyli do konkretnego, „fizycznego” adresu, gdzie te dane się mieszczą), a za pośrednictwem abstrakcyjnych „odwołań”, które mogą kierować w dowolne miejsce na świecie – albo w wiele takich miejsc. Plik przechowywany „w chmurze” nie jest już więc przywiązany do określonego miejsca – może być przetrzucany z miejsca na miejsce, rozdzielony na osobne segmenty albo przechowywany w wielu kopiach. W praktyce jednak, prędzej czy później, odwołanie takie będzie i tak prowadziło do jakiegoś konkretnego dysku w jakimś konkretnym budynku – co realnie, gdy mówimy o dużych zbiorach danych, oznacza właśnie tego typu centra danych.

[39*] Przedrostki do „tera” trochę już się upowszechniły, ale kolejne pamięta niewiele osób. Tymczasem coraz częściej będziemy mieli z nimi do czynienia. Proponuję prostą sztuczkę pomagającą zapamiętać kolejność następujących przedrostków. Wystarczy zwrócić uwagę na pierwsze litery skrótów od „peta” (P), „eksa” (E), „zetta” (Z) i „jotta” (Y), które razem z „tera” tworzą „TERA-P-E-Z-Y”, czyli prawie jak „trapezy”. Mnie pomagają.

[40*] W niektórych przypadkach zamiast mnożenia przez 1000 przyjmuje się mnożenie przez 1024 – w świecie komputerów króluje system dwójkowy, a 1024 to 2^{10} . Może to prowadzić do nieporozumień, ponieważ zdarza się, że, przykładowo, „megabajt”, który zgodnie z międzynarodowym układem SI powinien oznaczać $1000 \cdot 1000$, czyli 1 000 000 bajtów, używany jest na określenie $1024 \cdot 1024$, czyli 1 048 576 bajtów. Na szczęście dla potrzeb niniejszego rozdziału nie ma to większego znaczenia – mówimy i tak tylko o rządach wielkości, a 1000 i 1024 są szczęśliwie dość bliskie sobie.

[41*] Chociaż w 2015 roku dziennikarze pisali, że jest ich 40. Nie ma pełnej jawności na ten temat.

[42*] Na forach dyskusyjnych spekuluje się czasem, jakie standardy „pakowania” szaf serwerowych mogą stosować Chińczycy, jednak jest to w dużym stopniu wróżenie z fusów.

[43*] Nawet gigantyczne projekty typu Starlink nie zmieniają znacząco tego bilansu sił. W opublikowanym w kwietniu 2021 roku raporcie na temat internetu satelitarnego^[68] oszacowano, że łączny transfer danych oferowany przez wszystkich dostawców satelitarnych razem wziętych to 3 Tbps (terabity na sekundę). Pojedynczy kabel podmorski potrafi tymczasem przesyłać nawet kilkaset Tbps; kabli takich jest obecnie około 430 – zob. ramkę na kolejnej stronie.

[44*] Strat tych nie da się jednak całkowicie wyeliminować i w kablach o długości wielu kilometrów stosuje się „wzmacniacze”.

[45*] Na mapach Atlantyku przedstawia się 13 nazwanych kabli podmorskich, ale za pięcioma z tych nazw kryją się w istocie dwie niezależne nitki światłowodowe, tak więc w rzeczywistości ocean ten przekracza 18 fizycznie odrębnych kabli.

[46*] Choć, co ciekawe, wciąż istnieje spore i do tego rosące zapotrzebowanie na taśmy magnetyczne. Wielką zaletą zapisu magnetycznego jest bezpieczeństwo danych: jeżeli dane znajdują się na taśmie nawiniętej na szpulę, to ich odczytanie wymaga jej czysto fizycznego *przewinięcia*. Szpula z taśmą magnetyczną leżąca grzecznie w szafie to forma zapisu doskonale odporna na atak hakerski.

[47*] W momencie pisania tych słów rekord należy do bieżącej wersji strzelanki Call of Duty, „ważącej” 231 GB w wersji na PC.

[48*] Co więcej, lwią część przestrzeni dyskowej zajmują pewnie „nudne”, powtarzalne dane jak pliki systemu operacyjnego i popularne aplikacje, na przykład pakiety biurowe.

[49*] Liczba smartfonów na świecie przegoniła parę lat temu liczbę komputerów osobistych, ale średnia ilość pamięci w smartfonie ledwo przekracza 100 GB – wciąż więc większość „pamięci lokalnej” rezyduje w „pecetach”, choć smartfony bardzo szybko gonią „pecety” pod względem ilości pamięci, a już teraz prawdopodobnie przekraczają je ze względu na ilość *generowanych* w nich danych – o czym mówię niżej.

[50*] Co ciekawe, choć treści i usługi dostępne poprzez darknet bywają nielegalne, samo korzystanie z tej sieci i rozpowszechnianie adresów .onion takie nie jest, co prowadzi do zabawnej sytuacji: w pewnym sensie łatwiej uzyskać dostęp do *dark webu* niż do *deep webu*.

[51*] W jednym z badań darknetu oszacowano, że nieco powyżej połowy ruchu sieciowego w sieci Tor związana jest z aktywnością nielegalną^[70].

[52*] W rzeczywistości nazwa kategorii to: „HPAWVC” (Hack/Phreak/Anarchy/Warez/Virus/Crack). Mówiąc ogólnie, wygląda na to, że społeczność „hakerów” darknetu zdominowana jest przez względnie młode osoby, odgrywające rozmaite romantyczne fantazje, a niekoniecznie przez, hm, solidnych profesjonalistów.

[53*] Był to jeden z tych tematów, które przy projektowaniu tej książki odłożyłem do przegródki „nie” z największym bólem, ponieważ niemal na pewno odegrają swoją rolę w rozwoju technologii i społeczeństw w najbliższych latach. Nie da się jednak (chyba) napisać książki o wszystkim.

[54*] Pokrewnym, a nieco starszym pojęciem niż „rdzeń” jest „szkielet Internetu” (*Internet backbone*), obejmujący najważniejsze kable i „stacje przesyłowe”, łączące ze sobą kraje i kontynenty. Jeżeli parę stron wcześniej zabawiliście się w „hakera” i wykonaliście polecenie „tracert”, to ujawnił wam się fragment lokalnej mapy tego właśnie szkieletu. Nad infrastrukturą tą czuwają głównie duże firmy telekomunikacyjne.

[55*] To pięciu największych usługodawców „usług chmurowych” w 2019 roku.

[56*] Istniała swego czasu interesująca strona – Incognitube – prezentująca losowo wybrany film na YouTube, który miał mniej niż 100 wyświetleń. Spędziłem kiedyś sporo czasu, odświeżając tę stronę, zafascynowany niewyobrażalną liczbą amatorskich filmów przedstawiających bawiące się dzieci, krótkie klipy z wakacji oraz niezliczone nagrania ekranu komputerowego osób grających w strzelanki i Minecrafta. Większość z nich odnotowała mniej niż 10 wyświetleń.

[57*] Technicznie, w przypadku 5G odpowiednia technologia nosi nazwę „massive MIMO” (MIMO to „wiele wejść, wiele wyjść”). Tom Marzetta, współtwórca mMIMO, twierdzi, że rozszerzenie liczby urządzeń mogących naraz korzystać z sieci to kluczowa rzecz, nad którą pracuje w kontekście rozwoju 6G^[74].

[58*] Skale te rzeczywiście zaczynają się do siebie zbliżać. Zwykły magnetyczny dysk twardej z „talerzem” (czyli typu HDD) też potrzebuje chwili, aby „sięgnąć” do danych. Latencje dysków HDD podaje się w milisekundach i jest ona zwykle zbliżona do około 10 ms. Dużo szybsze są dyski typu SSD, w których nie znajduje się wirujący talerz i wędrująca głowica – latencja takiego dysku liczona jest już w mikrosekundach i wynosi około 100 μ s (0,1 ms). Cytowany wyżej raport IDC podaje, że udział dysków SSD w całym rynku dysków twardej szybko rośnie.

[59*] Inni mówią: Call Data Record.

[60*] Zgodnie z polskim prawem (art. 180 Prawa telekomunikacyjnego) operatorzy publicznych sieci telekomunikacyjnych mają obowiązek przechowywać dane o połączeniach przez 12 miesięcy i na życzenie udostępniać je uprawnionym podmiotom. Lista tych danych praktycznie pokrywa się z zakresem CDR.

[61*] To moje ostrożne oszacowanie na podstawie wielu różnych statystyk, które można znaleźć na ten temat. Parę przykładów. Firma analityczna Zenith szacuje, że przeciętny dorosły Amerykanin spędził w 2019 roku 3,5 godziny przy smartfonie. Twórcy aplikacji RescueTime twierdzą, że przeciętna osoba sięga po telefon 58 razy dziennie, przy czym w większości przypadków jest on używany krócej niż 2 minuty. W 2018 w czasopiśmie „Mobile Media & Communication” oszacowano^[75], na podstawie badań z użyciem specjalnie w tym celu zainstalowanej aplikacji monitorującej, że przeciętny amerykański student spędza 2 godziny i 40 minut przy smartfonie, w międzyczasie 101 razy przełączając się pomiędzy różnymi aplikacjami. Podobną metodą w 2019 roku oszacowano^[76], że średni czas oddawania się pojedynczej aktywności w czasie korzystania ze smartfona to 20 sekund.

[62*] Do takiego pakietu dołączane też było oszacowanie stopnia pewności. Oto fragment transmisji przechwyconej przez program monitorujący ruch sieciowy: "activities": [{"confidence": 99, "type": "onFoot"}, {"confidence": 99, "type": "walking"}].

[63*] Cytowany tu raport zawiera dość szczegółowy przepis pokazujący, w jaki sposób Google może „odanonimizować” wszelkie tego typu dane, opierając się na rozmaitych identyfikatorach dołączonych do wysyłanych przez smartfony raportów.

[64*] W maju 2020 roku ukazała się dość sucha, ale szczegółowa „przeglądówka” tych badań^[86].

[65*] Zwróćmy przy okazji uwagę, jak kluczowy jest pierwszy rok życia – głównie ze względu na powikłania okołoporodowe i skutki wad wrodzonych. Po przeżyciu pierwszego roku życia „tylko” 18 dzieci na 1000 umrze w ciągu następnych 14 lat. Statystyki międzynarodowe i polskie potwierdzają, że śmierć w wieku 1–15 lat jest *naprawdę* rzadka. To najzdrowszy i najbezpieczniejszy okres życia.

[66*] Zwyczajowo wyróżnia się trzy fazy badań klinicznych, każda kolejna prowadzona jest na większej grupie pacjentów i z coraz wyższymi oczekiwaniami. Ukończone z sukcesem badanie III fazy może być już podstawą do dopuszczenia danego leku albo procedury w terapii zwykłych pacjentów.

[67*] W poprzednim rozdziale trochę mówiliśmy o mikrometrach, czyli milionowych częściach metra; teraz schodzimy głębiej. Dla utrwalenia: 1000 nanometrów = 1 mikrometr, a 1000 mikrometrów = 1 milimetr (czyli tysięczna część metra).

[68*] Przynajmniej jeżeli zdefiniujemy problem jako „klasyczne” operowanie materia, tj. „bawienie się atomami”. Nikt nam nie zabroni marzyć o technologii w skali subatomowej. Odpowiednie słowa, analogiczne dla *nanotechnology*, są już w użyciu, przynajmniej w świecie wizjonerów i literatów: *picotechnology* dla hipotetycznej technologii w skali tysięcznych części nanometra, oraz *femtototechnology*, w skali jeszcze tysiąc razy mniejszej.

[69*] Nawiasem mówiąc, warto przypomnieć sobie w tym momencie narzekania twórców „Chemputera” – zob. ramka na s. 86 – że opisy chemików są niejednoznaczne. No bo co to właściwie znaczy „powoli” i po czym poznać, że bąbelki już „niemal zniknęły”?

[70*] Mówiąc technicznie, jest to więc kopolimer zbudowany z trzech typów monomerów. Jego formalna nazwa to „PMAEFc-ONB-PDMAEMA”. Smacznego!

[71*] Od angielskich fraz *double strand* (podwójna nić) i *single strand* (pojedyncza nić) pochodzą skrótowe określenia dsDNA i ssDNA, a także – w przypadku bliskiego kuzyna DNA – dsRNA i ssRNA. Skrótów te pojawiają się obficie w literaturze na temat biologii molekularnej i nanotechnologii, ale też w wirusologii. Kiedy przychodzi do opisanego jakiegoś wirusa, jedną z pierwszych informacji, jakie podaje się na jego temat, jest forma, w jakiej przechowuje swój materiał genetyczny – stąd cztery główne klasy wirusów: dsDNA, ssDNA, dsRNA i ssRNA.

[72*] A właściwie: sekwencję komplementarną wobec sekwencji tego genu – bo przecież chcemy, żeby nasza niteczka skleiła się z wybranym fragmentem naszego DNA. Przykładowo, jeżeli „celuję” w miejsce w moim genomie, w którym znajduje się sekwencja „GAGACCAATTTT”, to powinienem wyposażyć swojego nanorobota w sekwencję „CTCTGGTTAAAA”. Zachęcam do sprawdzenia literka po literce. Pamiętajmy, że pary komplementarnych podjednostek to A-T i C-G.

[73*] W momencie gdy pisałem pierwszą wersję tego rozdziału (czerwiec 2021), Google informował, że słowo „nanobiorobotics” występuje w internecie na zaledwie 105 stronach. W połowie lipca, gdy robiłem ostatnią korektę, było to już 135 wyników.

[74*] Cóż, na tym chyba polega cały dowcip. Coraz to bardziej aktywne i „inteligentne” cząsteczki leków w pewnym momencie stają się nanocząstkami, a potem nanorobotami – bez przekroczenia żadnej wyraźnej granicy, kiedy to z nieba pada strumień światła i dobiega tubalny głos: „Od tej pory nie jesteś już cząstką, jesteś

MASZYNA”. Na tej samej zasadzie nie ma momentu, w którym głos z nieba obwieszczałby, że „kupka głupich algorytmów” stała się właśnie „prawdziwą sztuczną inteligencją” – podobnie jak nie było takiego momentu, w którym prymitywny układ nerwowy nagle stał się „prawdziwym mózgiem”, a małpa stała się człowiekiem. Ewolucja, czy to naturalna, czy technologiczna, działa stopniowo, a zapaleni zwolennicy ostrych etykietek klasyfikacyjnych, próbujących powkładać pełne bogactwo świata do szufladek, zawsze przy tej okazji uprawiają trochę dopychania kolanem i pracowitego gumkowania etykiet.

[75*] Pierwszym autorem tego artykułu był Clyde Hutchinson, ale honorowe ostatnie miejsce na liście autorów zajmuje Craig Venter – i to nazwisko zdecydowanie warto zapamiętać na przyszłość. Jeżeli w najbliższych latach wydarzy się coś naprawdę spektakularnego w biologii syntetycznej, to jest spora szansa, że będzie to w laboratorium Ventera.

[76*] Są nieliczne wyjątki, jak choćby czerwona krwinka, która nie ma jądra komórkowego, a więc i materiału genetycznego.

[77*] *Tak naprawdę* jest to 23 pary chromosomów, z których każdy odziedziczyliśmy po jednym z rodziców. Są więc dwie kopie chromosomu nr 1, dwie kopie chromosomu nr 2... a na koniec dwa chromosomy płciowe: zestaw XY u mężczyzn i XX u kobiet. Biolog czytający tę książkę mógł więc poczuć ukłucie niepokoju, gdy padły słowa, że w każdej komórce znajduje się „jeden komplet” genomu, podczas gdy w zwykłej praktyce biologicznej mówi się o owych 46 chromosomach jako o dwóch kompletach, zaś zestaw obecny w plemniku i komórce jajowej, liczący 23 chromosomy, to dopiero „jeden komplet”. Jest dużo tego typu subtelności, z którymi jakoś muszę sobie radzić, jednocześnie nie spowalniając narracji – tym razem padło na przypis.

[78*] I znowu przypis. Pierwsza subtelność: zanim z DNA powstanie białko, najpierw tworzony jest „pośrednik” o nazwie RNA. Czasem jednak RNA takie nie jest wcale tłumaczone na białko i pozostaje po prostu cząstką RNA, pełniąc zupełnie inną funkcję niż skromny pośrednik. Tego typu cząstek RNA jest, jak się okazuje, całkiem sporo i pełnią one mnóstwo fascynujących funkcji w komórkach żywych. O fragmentach genomu, z których powstają takie cząstki, też zwyczajowo mówi się „geny”, chociaż niektórzy puryści zachęcają do twardego postawienia granicy: słowo „gen” stosujemy tylko wtedy, jeśli powstanie z niego białko. Nie mnie decydować, co jest słuszne; ja tu tylko relacjonuję. Druga „subtelność” jest taka, że fragmenty kodujące, czy to białka, czy RNA, zajmują tylko kilka procent genomu człowieka. Reszta to skomplikowany galimatias elementów regulujących albo pełniących funkcje „strukturalne” dla chromosomów, rozmaitych „powtórek” i innych segmentów, czasem o nie do końca jasnej funkcji.

[79*] Chociaż uczono mnie w szkole, że kolor oczu to jeden z nielicznych przykładów prostej zależności genetycznej opartej na jednym genie, to okazuje się, że nawet to nam odebrali źli naukowcy. Jest przynajmniej 15 genów odpowiadających za kolor oczu i niemal wszystkie kombinacje kolorów oczu rodziców i dzieci są dozwolone – choć niektóre są znacznie częstsze od innych. Krótko mówiąc, kiedy mówimy o dużych, ogólnych, widocznych gołym okiem cechach organizmu, to rzadko jest po prostu „gen na X” – najczęściej jest to raczej „zbiór genów wpływających na X”.

[80*] W literaturze naukowej zwykle stosuje się bardziej precyzyjne określenia, odwołujące się do etapu rozwoju embrionalnego, który jest w danym przypadku punktem wyjścia. Stąd na przykład „gastruloid” będący modelem gastruli.

[81*] Kojarzycie *Stayin' Alive* Bee Gees z filmu *Gorączka sobotniej nocy*? Rytm tej piosenki ma „beat” zbliżony do 2 Hz (2 razy na sekundę). Nawiasem mówiąc, w 2012 roku brytyjska organizacja British Heart Foundation zaleciła, aby osoby wykonujące ratunkowy masaż

serca uciskały klatkę piersiową w rytm tej piosenki. Rekomendację tę powtórzyły później również inne organizacje medyczne.

[82*] Elektroencefalogram – urządzenie do badania aktywności elektrycznej mózgu, najczęściej w postaci „czepka” z elektrodami przyklejanymi do głowy. O EEG będziemy sporo mówić w rozdziale V.

[83*] Technicznie, jest to więc nie tyle „lek”, co „prolek” – tak określa się substancję, która sama w sobie nie wykazuje pożądanej właściwości – dopiero po zajściu określonych reakcji chemicznych w organizmie człowieka powstaje inny związek, który jest właściwym lekiem.

[84*] Przypis dla pedantów: potoczne słowo „bakterie” należałoby zastąpić precyzyjniejszym słowem „prokarioty”, obejmującym dwie duże grupy jednokomórkowców: bakterie i archeany. Wspominam o tym, ponieważ odkrycie mechanizmu CRISPR/Cas nastąpiło najpierw właśnie u archeanów, a dopiero później u „bakterii w sensie ścisłym” (eubakterii). W dalszej części tekstu, który nie jest mimo wszystko podręcznikiem bakteriologii, będę jednak pisał po prostu „bakterie”.

[85*] Zgodnie z chińskim systemem zapisu imion i nazwisk, nazwisko rodowe – „He” – jest na pierwszym miejscu.

[86*] Nie ma żadnych wiarygodnych informacji o losie tych dziewczynek po wystąpieniu He 28 listopada 2018 roku. Co ciekawe, ze slajdów udostępnionych przez He wynika, że jedna z dwóch dziewczynek nie przeszła w pełni skutecznej modyfikacji genomu (nosi jedną kopię genu CCR5 zmodyfikowaną i jedną niezmodyfikowaną), a ponadto, że podobny eksperyment przeprowadzono na trzecim dziecku, które wówczas jeszcze się nie narodziło. Również i w temacie trzeciego „dziecka CRISPR” nie ma żadnych dodatkowych informacji poza spekulacjami.

[87*] Wynik jest dobry, biorąc pod uwagę skuteczność metod dostępnych obecnie biotechnologom, ale nie jest oczywiście idealny. Przypomina nam to przy okazji, dlaczego każda godzina rozwoju embrionalnego oddala nas od wizji „naprawienia genetycznego” człowieka. Jeżeli pracujemy z zygotą, czyli pojedynczą, pierwszą komórką organizmu, to wiemy, że cały organizm będzie nosił kopię jej genomu (choć może być też tak, że nie wszystkie spośród zaplanowanych zmian zostaną zrealizowane – skuteczność nie wynosi 100% zarówno ze względu na odsetek zmodyfikowanych komórek, jak i ze względu na odsetek wprowadzonych edycji w jednej komórce). Embrion tygodniowy ma tymczasem już około 500 komórek – w przypadku zapłodnienia naturalnego mało która kobieta ma zaś szansę wiedzieć, że nosi w sobie embrion w takim wieku. Nawet 90% skuteczność edycji genomu takiego embrionu oznacza, że będzie on nosił 50 komórek z nieusuniętymi wadami genetycznymi; z tych 50 komórek mogą się ostatecznie rozwinąć całe organy dorosłego organizmu.

[88*] W gronie tym znajduje się też polska tenisistka stołowa Natalia Partyka, urodzona bez prawej dłoni i przedramienia.

[89*] Tych z was, którzy zaczynają fantazjować o wymienieniu wszystkich swoich kości na tytanowe zamienniki, aby stać się niezniszczalnym cyborgiem, odsyłam do ramki na stronie 173, na której przypominam, że krążące w naszym ciele krwinki powstają w szpiku kostnym. Kości to znacznie więcej niż tylko „szkielet”!

[90*] Wszystkie podane tu wielkości są przybliżone. Każdy z wymienionych materiałów występuje w setkach odmian, więc teoretycznie trzeba by zawsze podawać, jaka konkretnie guma, jakie włókno nylonowe, jaka kość... i tak dalej. Nie przywiązuje się więc zbyt do konkretnych liczb. Sama kolejność w „rankingu” jest OK.

[91*] Dla zainteresowanych: istnieje pewne rzadko występujące zaburzenie psychiczne, wywołujące „niezgodę” na budowę własnego ciała, które w pewnych przypadkach prowadzi do przemożnej chęci poddania się amputacji. Chorobie tej towarzyszy poczucie wyobcowania z pewnego segmentu ciała – do tego stopnia, że niektórzy cierpiący na nią potrafią narysować długopisem na skórze granicę między ciałem „swoim” i „obcym”. Dziś zaburzenie to klasyfikowane jest jako dysforia integralności ciała (*body integrity dysphoria*). Z reguły nie udaje się znaleźć lekarza, który zgodziłby się na przeprowadzenie amputacji zdrowej kończyny, chociaż szkocki lekarz Robert Smith w latach 90. wykonał parę takich operacji, kontrowersyjnie twierdząc, że jest to właściwa metoda leczenia tego zaburzenia. W 2003 roku powstał na ten temat bardzo ciekawy dokument pt. *Whole* (reż. Melody Gilbert).

[92*] Przymiotnik „poznawczy”, zwykle stosowany jako tłumaczenie angielskiego *cognitive*, w ostatnich dekadach upowszechnił się jako określenie na wszystkie „względnie wysokie” funkcje mózgu ludzkiego. Od słowa *cognitive* pochodzi określenie *cognitive science*, czyli „nauki kognitywne” albo „kognitywistyka”. „Procesy poznawcze” to nieco więcej niż „procesy myślowe” (a tym bardziej „procesy świadome”) – przykładowo, proces uczenia się jazdy na rowerze jest procesem poznawczym, ale nie jest po prostu procesem myślowym (pojawiające się w trakcie nauki myśli typu „oho, jadę!” albo „muszę chyba pedałowac nieco szybciej” – to nie wszystko, co dzieje się w mózgu w trakcie nauki jazdy). Z drugiej strony w mózgu zachodzi wiele procesów, które nie są zwykle uważane za „poznawcze”, jak choćby regulowanie apetytu czy temperatury naszego ciała. Nie ma żadnej „oficjalnej” definicji tego terminu, ale czułem, że tego typu ogólne wyjaśnienie, nawet jeśli nie wszyscy kognitywiści czytający te słowa się z nim zgodzą, może się przydać na wstępie tego rozdziału.

[93*] Zdarzyło wam się kiedyś spojrzeć na dolny prawy róg czytanej właśnie najzupełniej papierowej książki, aby sprawdzić godzinę? Mnie tak. Mój mózg najwyraźniej przyzwyczał się już trochę, że informacja o czasie znajduje się w prawym dolnym rogu wielkiego białego obiektu z literkami – tak często czytam na laptopie. Widziałem też małe dzieci próbujące dwoma palcami „powiększyć” jakiś obrazek w gazecie i wykonujące gest „swipe’owania”, aby przewrócić stronę – co zabawne, ten drugi w zasadzie działa.

[94*] Tym terminem w elektronice obejmuje się element układu, który pozostaje w kontakcie z jego środowiskiem, odbierając z niego ładunki elektryczne lub je wysyłając. W kontekście medycyny owym „środowiskiem” jest ciało ludzkie – elektroda oznacza więc, w uproszczeniu, „nagi kabelek” zetknięty z ciałem tak, że aktywność elektryczna organizmu staje się przedłużeniem aktywności elektrycznej w obwodzie elektrycznym – i nawzajem. Obwód o budowie: „kabel miedziany–elektroda–nerw–elektroda–kabel miedziany” jest po prostu jednym obwodem elektrycznym.

[95*] Patrząc od zewnątrz, na drodze do mózgu znajdują się następujące warstwy: skóra, kość czaszki i opony (opona twarda, opona pajęczna (pajęczynówka) i opona miękka; pod pajęczynówką znajduje się też przestrzeń wypełniona płynem mózgowo-rdzeniowym). Cieniutka opona miękka jest już bezpośrednio zrosnięta z tkanką mózgu.

[96*] Rozmieszczenie tych 21 elektrod na powierzchni czaszki opisuje międzynarodowy standard o nazwie „10–20”, opisany w 1958 roku przez Międzynarodową Federację Neurofizjologii Klinicznej (International Federation of Clinical Neurophysiology, IFCN).

[97*] W ostatnich latach upowszechnia się stosowanie głębokiego uczenia maszynowego do „podkreśniania” rozdzielczości EEG, czasem ze spektakularnymi rezultatami[141].

[98*] Tytuł świetnej książki Jürgena Thorwalda o historii chirurgii mózgu. Gorąco polecam.

[99*] Jest też inne techniczne określenie dla ECoG – „wewnątrzczaszkowa elektroencefalografia” (iEEG) – które jeszcze wyraźniej pokazuje, że elektrokortykografia faktycznie jest odmianą EEG. Jest ono jednak rzadziej stosowane w literaturze.

[100*] Syndrom zamknięcia nie występuje wyłącznie wskutek ALS – może też zostać wywołany na przykład przez wylew krwi do pnia mózgu.

[101*] Mężczyzna ten cierpi na paraliż czterokończynowy po przerwaniu rdzenia między kręgami C4 i C5. W momencie powstawania artykułu wykazywał minimalną ruchliwość tylko w bicepsach, łokciu i lewym nadgarstku; wszystkie pozostałe mięśnie były całkowicie sparaliżowane.

[102*] Istnieje szereg technologii pozwalających na zasilanie bezprzewodowe, testowanych właściwie, od kiedy tylko zaczęły powstawać urządzenia elektryczne. Choć przyzwyczailiśmy się już, że informację można przekazywać „przez powietrze”, wciąż często zakładamy, że energia musi docierać do naszych urządzeń kablami – a przecież niekiedy nie! W ostatnich latach pewną popularność zaczyna zyskiwać tzw. ładowanie indukcyjne: urządzenie ładujące generuje pole magnetyczne, które jest wychwytywane przez ładowane urządzenie (na przykład telefon komórkowy). O ile ładowanie telefonu bez przewodu jest wyłącznie kwestią wygody, to w przypadku implantów jest to sprawa o znaczeniu kluczowym.

[103*] Szczególnie popularnym zamiennikiem dla palców są przysawki, nieustannie trwają jednak prace nad innymi rozwiązaniami. W maju 2021 roku opisano ciekawą alternatywę dla przysawek próżniowych (czyli takich, które tworzą szczelne połączenie z powierzchnią, a następnie odciągają z powstałej w ten sposób przestrzeni powietrze). Okazuje się, że zbliżony do płaskiej powierzchni elastyczny dysk, wibrujący z częstotliwością 200 Hz, całkiem sprawnie generuje siłę „przysysającą”^[149]. Ciekawym zbiegiem okoliczności dosłownie dzień później Toyota Research Institute upublicznił wyniki swoich prac nad jeszcze innym projektem: „chwytakiem bąbelkowym” o nazwie Puno Bubble Gripper, który wygląda trochę tak, jak dłoń z dwoma „palcami” w kształcie pałek, na powierzchni których znajdują się nadmuchiwane gumowe półkule. Przedmiot jest miękko przytrzymywany, jak gdyby trafił pomiędzy dwa baloniki.

[104*] Czysto technicznie, był to fragment tzw. kory przedruchowej – jest to obszar biorący udział w przygotowaniu ruchu.

[105*] „Etykieta nakazuje zgodność z istniejącymi zasadami”.

[106*] Ze względu na czułość dzisiejszych sejsmometrów nie musi być konieczne długie oczekiwanie. Ot, w trakcie pisania tego akapitu zarejestrowane zostało trzęsienie ziemi na Fidzi o sile 4,4 w skali Richtera – czyli lekkie bujanie, które nie wyrządziło raczej nikomu żadnej szkody. Fanom sejsmologii polecam tę aktualizowaną na bieżąco globalną mapę wstrząsów: <https://earthquake.usgs.gov/earthquakes/map/>.

[107*] W tym przypadku na ekranie wyświetlany był dłuższy fragment tekstu, nad którym przesuwano się „okienko”, tak że w danym momencie widoczne było tylko jedno słowo. Co ciekawe, zupełnie niezależnie od wspomnianych tu fińskich naukowców, istnieje grupa osób promujących faktyczne czytanie tekstów „po jednym słowie”. W tym modelu na ekranie zawsze wyświetla się tylko jedno słowo, dużą czcionką, w tym samym miejscu. Użytkownik wpatruje się więc w jeden punkt na ekranie, w którym wyświetlają się kolejne słowa zadanego tekstu. Jest to z początku trochę męcząca, ale dość szybka i na pewno interesująca alternatywa dla „zwykłego” czytania. Polecam spróbować samemu. W chwili pisania tej książki działa na przykład strona One Word Reader (<http://onewordreader.com/>), na której można wkleić do okienka dowolny tekst i wybrać

sobie samemu tempo „migania” (domyślnie ustawione jest 500 słów na minutę, czyli nieco ponad 8 na sekundę). Amazon oferuje też służący do tego program Word Runner.

[108*] Autorzy posługują się terminem *relevance*, czyli „istotność”, ale w praktyce w ich badaniu ujawnia się też aspekt ciekawości/zainteresowania.

- [1] S. Oh, Y.H. Oh, *Understanding the preference of the elderly for companion robot design*, w: *International Conference on Applied Human Factors and Ergonomics*, Springer, Cham 2019 (lipiec), s. 92–103.
- [2] M. Chignoli i in., *The MIT Humanoid Robot: Design, Motion Planning, and Control for Acrobatic Behaviors*, arXiv preprint arXiv:2104.09025, 2021.
- [3] H. Admoni, B. Scassellati, *Social eye gaze in human-robot interaction: A review*, „Journal of Human-Robot Interaction” 6(1), 2017, s. 25–63.
- [4] D. Karreman i in., *What happens when a robot favors someone? How a tour guide robot uses gaze behavior to address multiple persons while storytelling about art*, W: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2013 (marzec), s. 157–158.
- [5] B. Görer, A.A. Salah, H.L. Akin, *An autonomous robotic exercise tutor for elderly people*, „Autonomous Robots” 41(3), 2017, s. 657–678.
- [6] J. Owsianik, *State of the Sexbot Market*, <https://futureofsex.net/robots/state-of-the-sexbot-market-the-worlds-best-sex-robot-and-ai-love-doll-companies/> (29.11.2020).
- [7] M.B. Mathur, D.B. Reichling, *Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley*, „Cognition” 146, 2016, s. 22–32.
- [8] Na przykład: M. Raibert i in., *Bigdog, the rough-terrain quadruped robot*, „IFAC Proceedings Volumes” 41(2), 2008, s. 10822–10825.
- [9] A. Turing, *Computing machinery and intelligence-AM Turing*, „Mind” 59(236), 1950, s. 433.
- [10] J. Lee i in., *Learning quadrupedal locomotion over challenging terrain*, „Science Robotics” 5(47), 2020, eabc5986.
- [11] E. Ackerman, *Why Robots Can't Be Counted On to Find Survivors in the Florida Building Collapse*, <https://spectrum.ieee.org/automaton/robotics/robotics-hardware/why-robots-cant-help-find-survivors-in-the-florida-building-collapse> (26.06.2021).
- [12] K.N. McGuire i in., *Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment*, „Science Robotics” 4(35), 2019, eaaw9710.
- [13] R. Dubé i in., *An online multi-robot SLAM system for 3D LiDARs*, W: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017 (wrzesień), s. 1004–1011.
- [14] G. Li, R. Ge, G. Loianno, *Cooperative Transportation of Cable Suspended Payloads with MAVs Using Monocular Vision and Inertial Sensing*, „IEEE Robotics and Automation Letters” 6(3), 2021, s. 5316–5323.
- [15] A. Kehoe, M. Cecotti, *Multiple Destroyers Were Swarmed by Mysterious „Drones” off California over Numerous Nights*, <https://www.thedrive.com/the-war-zone/39913/multiple-destroyers-were-swarmed-by-mysterious-drones-off-california-over-numerous-nights> (23.03.2021).
- [16] <https://www.youtube.com/watch?v=9CO6M2HsoIA>
- [17] P. Scharre, *Why You Shouldn't Fear „Slaughterbots”*, <https://spectrum.ieee.org/automaton/robotics/military-robots/why-you-shouldnt-fear-slaughterbots> (22.12.2017).

- [18] P. Drapela, *Fake news? Lethal effect of micro drones*, <https://www.ar.admin.ch/en/armasuisse-wissenschaft-und-technologie-w-t/home.detail.news.html/ar-internet/news-2018/news-w-t/lethalmicrodrones.html> (11.04.2018).
- [19] D. Reid, *A swarm of armed drones attacked a Russian military base in Syria*, <https://www.cnn.com/2018/01/11/swarm-of-armed-drones-attacks-russian-military-base-in-syria.html> (11.01.2018).
- [20] S. Russell i in., *Lethal Autonomous Weapons Exist; They Must Be Banned*, <https://spectrum.ieee.org/automaton/robotics/military-robots/lethal-autonomous-weapons-exist-they-must-be-banned> (16.06.2021).
- [21] L. Majumdar Roy Choudhury i in., *Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011)*, United Nations Security Council, 2019 (9 grudnia).
- [22] S. Li i in., *Particle robotics based on statistical mechanics of loosely coupled components*, „Nature” 567(7748), 2019, s. 361–365.
- [23] M. Dorigo i in., *Swarmoid: A novel concept for the study of heterogeneous robotic swarms*, „IEEE Robotics & Automation Magazine” 20(4), 2013, s. 60–71.
- [24] J. Weizenbaum, *Computer Power and Human Reason. From Judgment to Calculation*, Oliver Wendell Holmes Library Phillips Academy, W.H. Freeman, San Francisco 1976.
- [25] A.M. Turing, J. Haugeland, *Computing Machinery and Intelligence*, MIT Press, Cambridge 1950, s. 29–56.
- [26] I. Kotseruba, J.K. Tsotsos, *40 years of cognitive architectures: Core cognitive abilities and practical applications*, „Artificial Intelligence Review” 53(1), 2020, s. 17–94.
- [27] P.W. Schermerhorn i in., *DIARC: A Testbed for Natural Human-Robot Interaction*, „AAAI” 6, 2006 (styczeń), s. 1972–1973.
- [28] B. Oosterveld, L. Brusatin, M. Scheutz, *Two bots, one brain: Component sharing in cognitive robotic architectures*, W: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017 (marzec), s. 415–415.
- [29] C. Rizzi i in., *A situation-aware fear learning (SAFEL) model for robots*, „Neurocomputing” 221, 2017, s. 32–47.
- [30] J. Thomason i in., *Opportunistic active learning for grounding natural language descriptions*, W: *Conference on Robot Learning*, PMLR, 2017 (październik), s. 67–76.
- [31] H. Ravichandar i in., *Recent advances in robot learning from demonstration*, „Annual Review of Control, Robotics, and Autonomous Systems” 3, 2020, s. 297–330.
- [32] Y. Liu i in., *Imitation from observation: Learning to imitate behaviors from raw video via context translation*, W: *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018 (maj), s. 1118–1125.
- [33] E. Ackerman, *It's (Still) Really Hard for Robots to Autonomously Do Household Chores*, <https://spectrum.ieee.org/automaton/robotics/home-robots/hard-for-robots-autonomous-household-chores> (17.06.2021).
- [34] https://www.youtube.com/watch?v=pv_n9FQRoZQ
- [35] <https://www.psynews.org/forums/topic/70402-the-psychedelic-trance-in-its-entirety/>

- [36] H. Kodama, *Automatic method for fabricating a three-dimensional plastic model with photo-hardening polymer*, „Review of Scientific Instruments” 52(11), 1981, s. 1770–1773.
- [37] J.R. Tumbleston i in., *Continuous liquid interface production of 3D objects*, „Science” 347(6228), 2015, s. 1349–1352.
- [38] K.C. Datsiou i in., *Additive manufacturing of glass with laser powder bed fusion*, „Journal of the American Ceramic Society” 102(8), 2019, s. 4410–4414.
- [39] A. Awad i in., *3D printing: Principles and pharmaceutical applications of selective laser sintering*, „International Journal of Pharmaceutics” 586, 2020, 119594.
- [40] A. Behroozfar i in., *Microscale 3D printing of nanotwinned copper*, „Advanced Materials” 30(4), 2018, s. 1705107.
- [41] R. Matsuzaki, T. Kanatani, A. Todoroki, *Multi-material additive manufacturing of polymers and metals using fused filament fabrication and electroforming*, „Additive Manufacturing” 29, 2019, 100812.
- [42] M. Dey, I.T. Ozbolat, *3D bioprinting of cells, tissues and organs*, „Scientific Reports” 10, 2020, 14023.
- [43] https://reprap.org/wiki/Snappy_3
- [44] S. Steiner i in., *Organic synthesis in a modular robotic system driven by a chemical programming language*, „Science” 363(6423), 2019, eaav2211.
- [45] S. Chatterjee i in., *Automated radial synthesis of organic molecules*, „Nature” 579(7799), 2020, s. 379–384.
- [46] A.C. Vaucher i in., *Automated extraction of chemical synthesis actions from experimental procedures*, „Nature Communications” 11(1), 2020, s. 1–11.
- [47] <http://www.chem.gla.ac.uk/cronin/chemify/>
- [48] R. Lever, *Hacking the food chain, Silicon Valley style*, <https://phys.org/news/2015-07-hacking-food-chain-silicon-valley.html> (3.07.2015).
- [49] <https://www.media.mit.edu/projects/personal-food-computer/>
- [50] <https://www.opensourceecology.org/gvcs/>
- [51] T. Mikolov i in., *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781, 2013.
- [52] <http://wiki.pathmind.com/word2vec>
- [53] V. Tshitoyan i in., *Unsupervised word embeddings capture latent knowledge from materials science literature*, „Nature” 571(7763), 2019, s. 95–98.
- [54] A. Abid, M. Farooqi, J. Zou, *Large language models associate Muslims with violence*, „Nature Machine Intelligence” 3(6), 2021, s. 461–463.
- [55] N. Slonim i in., *An autonomous debating system*, „Nature” 591(7850), 2021, s. 379–384.
- [56] <https://github.com/lucidrains/big-sleep>
- [57] <https://softologyblog.wordpress.com/>
- [58] <https://rynmurdock.github.io/2021/02/26/Aleph2Image.html>
- [59] <https://openai.com/blog/dall-e/>

[60] <https://www.thispersondoesnotexist.com/>

[61] S. Suwajanakorn, S.M. Seitz, I. Kemelmacher-Shlizerman, *Synthesizing Obama: Learning lip sync from audio*, „ACM Transactions on Graphics (ToG)” 36(4), 2017, s. 1–13.

[62] <https://www.flawlessai.com/>

[63] H. Ahn i in., *Text2action: Generative adversarial synthesis from language to action*, W: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018 (maj), s. 5915–5920.

[64] <https://www.youtube.com/watch?v=wYb3Wimn01s>

[65] T.W. Tam, C. Ko, *China Data Centre Sector*, „DBS Asian Insights. Sector Briefing” 70.

[66] R. Danilak, *Why energy is a big and rapidly growing problem for data centers*, <https://www.forbes.com/sites/forbestechcouncil/2017/12/15/why-energy-is-a-big-and-rapidly-growing-problem-for-data-centers> (15.12.2017).

[67] <https://www.techrepublic.com/article/95-of-global-data-center-traffic-will-be-from-the-cloud-by-2021/>

[68] J. Garrity, A. Husar, *Digital Connectivity and Low Earth Orbit Satellite Constellations. Opportunities for Asia and the Pacific*, Asian Development Bank, Manila 2021.

[69] M.K. Bergman, *White paper: The deep Web: surfacing hidden value*, „Journal of electronic publishing” 7(1), 2001.

[70] D. Moore, T. Rid, *Cryptopolitik and the darknet*, „Survival: Global Politics and Strategy” 58, 2016, s. 7–38.

[71] EMCDDA, *Drugs and the darknet: perspectives for enforcement, research and policy*, EMCDDA, Europol, Lizbona 2017 (listopad).

[72] D. Reinsel, J. Gantz., J. Rydning, *The Digitization of the World from Edge to Core*, International Data Corporation, Framingham 2018.

[73] Cisco Annual Internet Report (2018–2023) White Paper, <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (9.03.2020).

[74] M. Koziol, *Here's What 6G Will Be, According to the Creator of Massive MIMO*, <https://spectrum.ieee.org/tech-talk/telecom/wireless/heres-what-6g-will-be-according-to-the-creator-of-massive-mimo> (08.06.2021).

[75] T. Deng i in., *Measuring smartphone usage and task switching with log tracking and self-reports*, „Mobile Media & Communication” 7(1), 2019, s. 3–23.

[76] B. Reeves i in., *Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them*, „Human–Computer Interaction”, 2019, s. 1–52.

[77] D. Schmidt, Google data collection, <https://digitalcontentnext.org/blog/2018/08/21/google-data-collection-research/> (21.08.2018).

[78] R. Nakashima, *AP Exclusive: Google tracks your movements, like it or not*, <https://apnews.com/828aefab64d4411bac257a07c1af0ecb> (13.08.2018).

[79] S. Perez, *TikTok just gave itself permission to collect biometric data on US users, including „faceprints and voiceprints”*, <https://techcrunch.com/2021/06/03/tiktok-just-gave-itself-permission-to-collect-biometric-data-on-u-s-users-including-faceprints-and-voiceprints/> (4.06.2021).

- [80] <https://www.tiktok.com/legal/privacy-policy?lang=en#privacy-us>
- [81] <https://www.socialmediatoday.com/news/instagram-adds-automatic-image-descriptions-to-assist-visually-impaired-use/543172/>
- [82] <https://ai.googleblog.com/2018/11/a-structured-approach-to-unsupervised.html>
- [83] K. Wiggers, *Facebook: Our video selfie test is for spotting bots, doesn't use facial recognition*, <https://venturebeat.com/2019/11/05/facebook-our-video-selfie-test-is-for-spotting-bots-doesnt-use-facial-recognition/> (5.11.2019).
- [84] <https://www.salon.com/2019/02/10/you-just-deleted-facebook-can-you-trust-facebook-to-delete-your-data/>
- [85] <https://faq.whatsapp.com/android/account-and-profile/how-to-delete-your-account/?lang=en>
- [86] H. Vora, M. Bhamare, D.K.A. Kumar, *Personality Prediction from Social Media Text: An Overview*, „International Journal of Engineering Research” 9(5), 2020, s. 352–357.
- [87] M.M. Tadesse, H. Lin, B. Xu, L. Yang, *Personality predictions based on user behavior on the Facebook social media platform*, „IEEE Access” 6, 2018, s. 61959–61969.
- [88] A. Laleh, R. Shahram, *Analyzing Facebook activities for personality recognition*, W: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2017 (grudzień), s. 960–964.
- [89] R.B. Correia, I.B. Wood, J. Bollen, L.M. Rocha, *Mining social media data for biomedical signals and health-related behavior*, „Annual Review of Biomedical Data Science” 3, 2020, s. 433–458.
- [90] J. Zomick, S.I. Levitan, M. Serper, *Linguistic analysis of schizophrenia in Reddit posts*, W: *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 2019 (czerwiec), s. 74–83.
- [91] R.B. Correia, L. Li, L.M. Rocha, *Monitoring potential drug interactions and reactions via network analysis of instagram user timelines*, W: *Biocomputing 2016: Proceedings of the Pacific Symposium*, 2016, s. 492–503.
- [92] I.B. Wood, P.L. Varela, J. Bollen, L.M. Rocha, J. Gonçalves-Sá, *Human sexual cycles are driven by culture and match collective moods*, „Scientific Reports” 7(1), 2017, s. 1–11.
- [93] G. Corfield, *Nine million logs of Brits' road journeys spill onto the internet from password-less number-plate camera dashboard*, https://www.theregister.com/2020/04/28/anpr_sheffield_council/ (28.04.2020).
- [94] J.C. Riley, *Estimates of regional and global life expectancy, 1800–2001*, „Population and Development Review” 31(3), 2005, s. 537–543.
- [95] J.W. Vaupel, F. Villavicencio, M.P. Bergeron-Boucher, *Demographic perspectives on the rise of longevity*, „Proceedings of the National Academy of Sciences” 118(9), 2021.
- [96] A.A. Volk, J.A. Atkinson, *Infant and child death in the human environment of evolutionary adaptation*, „Evolution and Human Behavior” 34(3), 2013, s. 182–192.
- [97] J.W. Vaupel, F. Villavicencio, M.P. Bergeron-Boucher, *Demographic perspectives on the rise of longevity*, „Proceedings of the National Academy of Sciences” 118(9), 2021.
- [98] M.C. Jin, Z.A. Medress, T.D. Azad, V.M. Doulames, A. Veeravagu, *Stem cell therapies for acute spinal cord injury in humans: A review*, „Neurosurgical Focus” 46(3), 2019, E10.

- [99] N. Tellisi, N.A. Ashammakhi, F. Billi, O. Kaarela, *Three dimensional printed bone implants in the clini*, „Journal of Craniofacial Surgery” 29(8), 2018, s. 2363–2367.
- [100] P. Macchiarini, *Clinical transplantation of a tissue-engineered airway*, „The Lancet” 372(9655), 2008, s. 2023–2030.
- [101] J. Willemse i in., *Fast, robust and effective decellularization of whole human livers using mild detergents and pressure controlled perfusion*, „Materials Science and Engineering: C” 108, 2020, 110200.
- [102] Y. Li i in., *Decellularization of porcine whole lung to obtain a clinical-scale bioengineered scaffold*, „Journal of Biomedical Materials Research Part A” 109(9), 2021, s. 1623–1632.
- [103] I. Pla-Palacín i in., *Liver bioengineering using decellularized whole-liver scaffolds*, W: *Decellularized Scaffolds and Organogenesis*, Humana Press, New York 2017, s. 293–305.
- [104] B.M. Wiebe, H. Laursen, *Human lung volume, alveolar surface area, and capillary length*, „Microscopy Research and Technique” 32(3), 1995, s. 255–262.
- [105] A.P. Golden, J. Tien, *Fabrication of microfluidic hydrogels using molded gelatin as a sacrificial element*, „Lab on a Chip” 7(6), 2007, s. 720–725.
- [106] E.A. Bulanova i in., *Bioprinting of a functional vascularized mouse thyroid gland construct*, „Biofabrication” 9(3), 2017, 034105.
- [107] M. Vert i in., *Terminology for biorelated polymers and applications (IUPAC Recommendations 2012)*, „Pure and Applied Chemistry” 84(2), 2012, s. 377–410.
- [108] K. Zhang i in., *Synthesis of temperature, pH, light and dual-redox quintuple-stimuli-responsive shell-crosslinked polymeric nanoparticles for controlled release*, „Materials Science and Engineering: C” 87, 2018, s. 1–9.
- [109] B.A. Badeau, C.A. DeForest, *Programming stimuli-responsive behavior into biomaterials*, „Annual Review of Biomedical Engineering” 21, 2019, s. 241–265.
- [110] R. Veneziano i in., *Designer nanoscale DNA assemblies programmed from the top down*, „Science” 352(6293), 2016, s. 1534–1534.
- [111] S. Dey i in., *DNA origami*, „Nature Reviews Methods Primers” 1(1), 2021, s. 1–24.
- [112] L. Yang i in., *An intelligent DNA nanorobot for autonomous anticoagulation*, „Angewandte Chemie” 132(40), 2020, s. 17850–17857.
- [113] F. Cao, H. Sato, *Remote radio controlled insect-computer hybrid legged robot*, W: *2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, IEEE, 2017 (czerwiec), s. 59–62.
- [114] F. Striggow i in., *Sperm-Driven Micromotors Moving in Oviduct Fluid and Viscoelastic Media*, „Small” 16(24), 2020, 2000213.
- [115] L.K. Abdelmohsen i in., *Dynamic loading and unloading of proteins in polymeric stomatocytes: Formation of an enzyme-loaded supramolecular nanomotor*, „ACS Nano” 10(2), 2016, s. 2652–2660.
- [116] P. Ge i in., *Action of a minimal contractile bactericidal nanomachine*, „Nature” 580(7805), 2020, s. 658–662.
- [117] O. Felfouli i in., *Magneto-aerotactic bacteria deliver drug-containing nanoliposomes to tumour hypoxic regions*, „Nature Nanotechnology” 11(11), 2016, s. 941–947.

- [118] C.A. Hutchison i in., *Design and synthesis of a minimal bacterial genome*, „Science” 351(6280), 2016, aad6253.
- [119] P. He i in., *The changing mouse embryo transcriptome at whole tissue and single-cell resolution*, „Nature” 583(7818), 2020, s. 760–767.
- [120] L.E. Byrnes i in., *Lineage dynamics of murine pancreatic development at single-cell resolution*, „Nature Communications” 9(1), 2018, s. 1–17.
- [121] K. Tiklová i in., *Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development*, „Nature Communications” 10(1), 2019, s. 1–12.
- [122] I. Martyn, E.D. Siggia, A.H. Brivanlou, *Mapping cell migrations and fates in a gastruloid model to the human primitive streak*, „Human Development” 146(17), 2019, dev179564.
- [123] G. Quadrato i in., *Cell diversity and network dynamics in photosensitive human brain organoids*, „Nature” 545(7652), 2017, s. 48–53.
- [124] C.A. Trujillo i in., *Complex oscillatory waves emerging from cortical organoids model early human brain network development*, „Cell. Stem Cell” 25(4), 2019, s. 558–569.
- [125] M. Lancaster, *Brain organoids*, <http://serious-science.org/brain-organoids-10510> (5.04.2021).
- [126] S.A.P. Rajan i in., *Probing prodrug metabolism and reciprocal toxicity with an integrated and humanized multi-tissue organ-on-a-chip platform*, „Acta Biomaterialia” 106, 2020, s. 124–135.
- [127] F. Zheng i in., *Patient-Specific Organoid and Organ-on-a-Chip: 3D Cell--Culture Meets 3D Printing and Numerical Simulation*, „Advanced Biology” 5(6), 2021, 2000024.
- [128] P. Liang i in., *CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes*, „Protein & Cell” 6(5), 2015, s. 363–372.
- [129] E. Lanphier i in., *Don't edit the human germ line*, „Nature News” 519(7544), 2015, s. 410.
- [130] A.V. Anzalone i in., *Search-and-replace genome editing without double-strand breaks or donor DNA*, „Nature” 576(7785), 2019, s. 149–157.
- [131] https://www.ted.com/talks/hugh_herr_the_new_bionics_that_let_us_run_climb_and_dance
- [132] https://www.ted.com/talks/hugh_herr_how_we_ll_become_cyborgs_and_extend_human_potential
- [133] H. Chu i in., *Unipolar stroke, electroosmotic pump carbon nanotube yarn muscles*, „Science” 371(6528), 2021, s. 494–498.
- [134] A. Takakura i in., *Strength of carbon nanotubes depends on their chemical structures*, „Nature Communications” 10(1), 2019, s. 1–7.
- [135] D. Nelidova i in., *Restoring light sensitivity using tunable near-infrared sensors*, „Science” 368(6495), 2020, s. 1108–1113.
- [136] K.G. Gawit i in., *Effect of cellular mobile phone use and cetirizine on hand-eye coordination and visual acuity*, „Journal of Clinical and Diagnostic Research” 11(9), 2017, FC09.
- [137] J.M. Lu, Y.C. Lo, *Investigation of smartphone use while walking and its influences on one's behavior among pedestrians in Taiwan*, W: *International Conference on Human-Computer Interaction*, Springer, Cham 2017 (lipiec), s. 469–475.

- [138] J. Marty-Dugas i in., *The relation between smartphone use and everyday inattention*. *Psychology of Consciousness: Theory, „Research, and Practice”* 5(1), 2018, s. 46.
- [139] G.W. Small i in., *Brain health consequences of digital technology use*, „Dialogues in Clinical Neuroscience” 22(2), 2020, s. 179.
- [140] A.F. Ward i in., *Brain drain: The mere presence of one’s own smartphone reduces available cognitive capacity*, „Journal of the Association for Consumer Research” 2(2), 2017, s. 140–154.
- [141] I.A. Corley, Y. Huang, *Deep EEG super-resolution: Upsampling EEG spatial resolution with generative adversarial networks*, W: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2018, s. 100–103.
- [142] S.K. Brodnick i in., *µECoG recordings through a thinned skull*, „Frontiers in Neuroscience” 13, 2019, 1017.
- [143] T.J. Oxley i in., *Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis: First in-human experience*, „Journal of Neurointerventional Surgery” 13(2), 2021, s. 102–108.
- [144] T. Kaiju i in., *High spatiotemporal resolution ECoG recording of somatosensory evoked potentials with flexible micro-electrode arrays*, „Frontiers in Neural Circuits” 11, 2017, 20.
- [145] K.J. Miller, D. Hermes, N.P. Staff, *The current state of electrocorticography-based brain-computer interfaces*, „Neurosurgical Focus” 49(1), 2020, E2.
- [146] S. Silvestro i in., *Stem cells therapy for spinal cord injury: An overview of clinical trials*, „International Journal of Molecular Sciences” 21(2), 2020, s. 659.
- [147] P.R. Kennedy, R.A. Bakay, *Restoration of neural output from a paralyzed patient by a direct brain connection*, „Neuroreport” 9(8), 1998, s. 1707–1711.
- [148] A.L. Benabid i in., *An exoskeleton controlled by an epidural wireless brain-machine interface in a tetraplegic patient: A proof-of-concept demonstration*, „The Lancet Neurology” 18(12), 2019, s. 1112–1122.
- [149] W.P. Weston-Dawkes i in., *Gas-Lubricated Vibration-Based Adhesion for Robotics*, „Advanced Intelligent Systems” 3(7), 2021, 2100001.
- [150] S.R. Kashef, S. Amini, A. Akbarzadeh, *Robotic hand: A review on linkage-driven finger mechanisms of prosthetic hands and evaluation of the performance criteria*, „Mechanism and Machine Theory” 145, 2020, 103677.
- [151] C. Pandarinath i in., *High performance communication by people with paralysis using an intracortical brain-computer interface*, „eLife” 6, 2017, e18554.
- [152] F.R. Willett i in., *High-performance brain-to-text communication via handwriting*, „Nature” 593(7858), 2021, s. 249–254.
- [153] G.K. Anumanchipalli, J. Chartier, E.F. Chang, *Speech synthesis from neural decoding of spoken sentences*, „Nature” 568(7753), 2019, s. 493–498.
- [154] M. Ortiz-Catalan i in., *Self-contained neuromusculoskeletal arm prostheses*, „New England Journal of Medicine” 382(18), 2020, s. 1732–1738.
- [155] T.H. Nguyen, W.Y. Chung, *Detection of driver braking intention using EEG signals during simulated driving*, „Sensors” 19(13), 2019, 2863.
- [156] N.P. Sarode, T.S. Lakshmi, V. Chavan, *A Brain Computer Interface for Smart Home Control*, „International Journal of Science, Engineering and Technology Research” 6(5),

2017, s. 918–922.

[157] X. Zhang i in., *Internet of Things meets brain-computer interface: A unified deep learning framework for enabling human-thing cognitive interactivity*, „IEEE Internet of Things Journal” 6(2), 2018, s. 2084–2092.

[158] S.M. Polyn i in., *Category-specific cortical activity precedes retrieval during memory search*, „Science” 310(5756), 2005, s. 1963–1966.

[159] K. Seeliger i in., *Generative adversarial networks for reconstructing natural images from brain activity*, „NeuroImage” 181, 2018, s. 775–785.

[160] G.M. McKhann i in., *Intraoperative hippocampal electrocorticography to predict the extent of hippocampal resection in temporal lobe epilepsy surgery*, „Journal of neurosurgery” 93(1), 2000, s. 44–52.

[161] M.J. Eugster i in., *Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals*, „Scientific Reports” 6(1), 2016, s. 1–10.

[162] L. Kangassalo, M. Spapé, T. Ruotsalo, *Neuroadaptive modelling for generating images matching perceptual categories*, „Scientific Reports” 10(1), 2020, s. 1–10.

[163] *Scientists Discover How to Upload Knowledge to Your Brain*, <https://www.scinature.vip/2021/06/scientists-discover-how-to-upload.html> (6.6.2021).