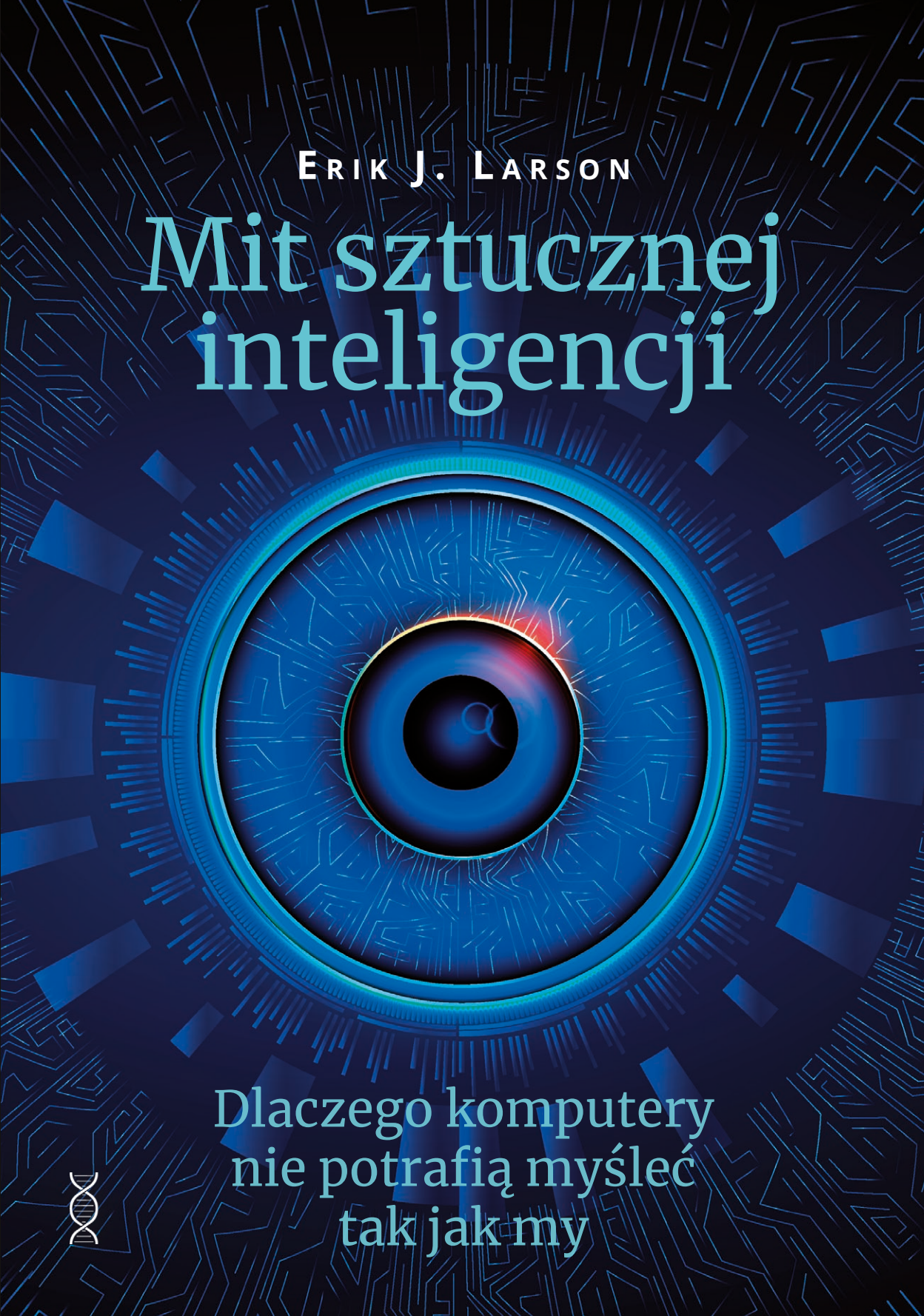
Erik J. Larson

Mit sztucznej inteligencji



Dlaczego komputery nie potrafią myśleć tak jak my

Mit sztucznej inteligencji

ERIK J. LARSON

Mit sztucznej inteligencji

Dlaczego komputery nie potrafią myśleć tak jak my

EN ARCHE
wszechświat człowiek nauka

Warszawa 2025

Tytuł oryginału
The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do

Przekład
Piotr Bylica

Redaktor prowadzący
Anna Kaszubowska

Redakcja językowa
Anna Kaszubowska

Korekta
Beata Gołkowska

Projekt okładki
Jadwiga Topolowska

Skład
Amadeusz Targoński

Wydanie I

ISBN 978-83-68119-16-9 (PDF)
ISBN 978-83-68119-15-2 (EPUB)

Fundacja En Arche
al. Niepodległości 124, lok. 26
02-577 Warszawa
biuro@enarche.pl
Księgarnia internetowa
enarche.pl/ksiegarnia/

Dla Brooke i Bena

Spis treści

Wprowadzenie

Na kartach tej książki przeczytacie o micie sztucznej inteligencji (AI). Mitem nie jest przekonanie, że prawdziwa AI jest możliwa. Co do tego przyszłość AI pozostaje naukową niewiadomą. Mówiąc o micie sztucznej inteligencji, mam na myśli przekonanie, że jej nadejście jest nieuniknione i to tylko kwestia czasu; że już wkroczyliśmy na drogę, która doprowadzi do AI będącej na poziomie inteligencji człowieka, a następnie do superinteligencji. Wcale tak nie jest. Ta droga istnieje tylko w naszej wyobraźni. Niemniej jednak przekonanie o nieuchronności AI jest tak głęboko zakorzenione w popularyzatorskim wymiarze dyskusji – promowanej przez ekspertów w mediach, liderów idei takich jak Elon Musk, a nawet wielu naukowców zajmujących się AI (choć na pewno nie wszystkich) – że argumentowanie przeciwko temu często bywa traktowane jako forma luddyzmu[1] lub przynajmniej jako wyraz krótkowzroczności wizji przyszłości technologii oraz jako potencjalne zagrożenie polegające na tym, że nie będziemy przygotowani do życia w świecie inteligentnych maszyn.

Jak pokażę, naukowe badania nad AI sprawiły, że zdaliśmy sobie sprawę, że jeśli chodzi o podstawy inteligencji, mamy do czynienia z tajemnicą i nikt obecnie nie wie, jak tę zagadkę rozwiązać. Zwolennicy idei o możliwości opracowania AI są silnie zmotywowani, aby minimalizować znaczenie jej znanych ograniczeń. W końcu branża AI to wielki biznes, a jej dominujące oddziaływanie na kulturę jest coraz większe. Jednak niezależnie od tego, czy nam się to podoba, czy nie, obecna wiedza o naturze inteligencji wskazuje, że możliwości przyszłych systemów AI są ograniczone. Powinniśmy to powiedzieć wprost: wszystkie świadectwa sugerują, że inteligencja ludzka i maszynowa są radykalnie

[1] Słowo „luddyzm" pochodzi od nazwiska prawdopodobnie wyimaginowanego angielskiego tkacza Neda Ludda, który około 1779 roku miał rzekomo zniszczyć krosna tkackie; termin odnosi się do idei wyznawanych przez jego zwolenników z początku XIX wieku, sprzeciwiających się użyciu maszyn włókienniczych w Anglii. Współczesny luddysta jest sceptyczny wobec każdej nowej technologii i sprzeciwia się wszelkim zmianom technologicznym, które uderzają w ludzką naturę i zagrażają realizacji potrzeb człowieka (przyp. tłum.).

różne. Mit AI głosi, że różnice te są tylko tymczasowe i potężniejsze systemy ostatecznie je zlikwidują. Futuryści tacy jak Ray Kurzweil oraz filozof Nick Bostrom, prominentni zwolennicy mitu, nie mówią tylko o tym, że osiągnięcie AI na poziomie ludzkim jest nieuniknione, ale twierdzą także, iż wkrótce po jej nadejściu superinteligentne maszyny zostawią nas daleko w tyle.

Ta książka wyjaśnia dwa ważne aspekty mitu o AI: naukowy i kulturowy. Naukowa część mitu zakłada, że wystarczy, abyśmy dalej „szli drogą" wyzwań związanych z ogólną inteligencją, dokonując postępów w wąskich zastosowaniach inteligencji, takich jak granie w gry czy rozpoznawanie obrazów. To poważny błąd, ponieważ sukces w wąskich zastosowaniach nie zbliża nas ani na krok do inteligencji ogólnej. Biorąc pod uwagę naszą obecną wiedzę o AI, wnioskowania, które musiałyby umieć przeprowadzać systemy ogólnej inteligencji – aby przeczytać gazetę, prowadzić podstawową rozmowę lub być pomocnym jak robot Rosie z *Jetsonów* – nie mogą być zaprogramowane, nauczone ani wytworzone. Podczas gdy skutecznie stosujemy prostsze, wyspecjalizowane wersje inteligencji, które korzystają z szybszych komputerów i dużej ilości danych, nie robimy stopniowych postępów na drodze do ogólnej sztucznej inteligencji, lecz raczej zbieramy najłatwiej dostępne owoce. Przeskok do ogólnego „zdrowego rozsądku" to zupełnie inna sprawa i nie znamy drogi łączącej jedno z drugim. Nie istnieje żaden algorytm dla ogólnej inteligencji. Mamy ponadto dobre powody dla sceptycyzmu co do tego, że taki algorytm wyłoni się w wyniku dalszych wysiłków na rzecz systemów uczenia głębokiego lub jakiegokolwiek innego podejścia, które jest dzisiaj popularne. Znacznie bardziej prawdopodobne jest, że konieczny będzie znaczący przełom naukowy, a nikt obecnie nie ma najmniejszego pojęcia, jak taki przełom mógłby wyglądać, nie mówiąc już o szczegółach tego, jak go osiągnąć.

Mit o AI jest zatem szkodliwy, ponieważ mówienie bez końca o stałych postępach ukrywa fakt znanej nauce tajemnicy. Mit podtrzymuje wiarę w nieuchronny sukces, ale prawdziwy szacunek dla nauki powinien skłonić nas do powrotu do punktu wyjścia. W ten sposób przechodzimy do drugiego tematu – kulturowych konsekwencji mitu. Kierowanie się tym mitem nie jest dobrym sposobem na bycie „dobrze poinformowanym przez źródła" ani nawet wyrazem przyjęcia postawy neutralnej. Jest szkodliwe dla nauki i dla nas samych. Dlaczego? Jednym z powodów jest to, że mało prawdopodobne jest opracowanie innowacji, jeśli zamiast stawić czoła rzeczywistości, zdecydujemy się zignorować centralną tajemnicę. W zdrowej kulturze innowacji podkreśla

się znaczenie eksploracji nieznanych obszarów; nie wspiera się przesadnie długiego trwania przy znanych metodach – szczególnie gdy te metody okazały się niewystarczające, by pchnąć nas znacznie dalej. Mitologia dotycząca nieuchronnego sukcesu w obszarze AI ma tendencję do psucia samej kultury wynalazczości niezbędnej do rzeczywistego postępu, czy to postępu z AI podobnej ludziom, czy bez niej. Mit ten również zachęca do biernego pogodzenia się z powolnym rozwojem świata maszyn, w którym prawdziwe wynalazki są spychane na margines na rzecz futurystycznych narracji promujących obecne podejścia, często wspieranych przez ugruntowane interesy.

Kto powinien przeczytać tę książkę? Z pewnością każdy, kto jest podekscytowany ideą AI, ale zastanawia się, dlaczego jej realizacja zawsze jest odległa o 10 lub 20 lat. Wyjaśnię, jaki naukowy powód się za tym kryje. Powinieneś również przeczytać tę książkę, jeśli uważasz, że postęp AI w kierunku superinteligencji jest nieunikniony i obawiasz się, co zrobić, gdy już nastąpi. Chociaż nie mogę udowodnić, że pewnego dnia AI nie przejmie władzy, mogę podać powody, by poważnie wątpić w prawdopodobieństwo takiego scenariusza. Ogólnie rzecz biorąc, powinieneś przeczytać tę książkę, jeśli jesteś po prostu ciekawy, a jednocześnie zdezorientowany powszechnym szałem na punkcie AI, z jakim mamy do czynienia w naszym społeczeństwie. Wyjaśnię pochodzenie mitu o AI, co wiemy i czego nie wiemy o perspektywach rzeczywistego osiągnięcia AI na poziomie ludzkim oraz dlaczego musimy lepiej docenić jedyną prawdziwą inteligencję, którą znamy – naszą własną.

Treść książki

W części pierwszej, zatytułowanej *Uproszczony świat*, wyjaśniam, jak w powszechnym podejściu do AI uproszczono znaczenie pojęć odnoszących się do ludzi, jednocześnie rozszerzając idee dotyczące technologii. Zaczęło się to od Alana Turinga, twórcy dziedziny AI, i obejmowało zrozumiałe, ale niefortunne uproszczenia, które nazywam „błędnymi koncepcjami inteligencji". Za sprawą przyjaciela Turinga i statystyka Irvinga Johna Goode'a, który wprowadził pojęcie „ultrainteligencji" mające odpowiadać przewidywanej konsekwencji osiągnięcia AI na poziomie ludzkim, początkowe błędy osiągnęły status ideologii. Przyglądając się temu, co głosili Turing i Good, widzimy, jak kształtował się nowoczesny mit AI. Jego rozwój wprowadził nas w erę, którą nazywam „technologicznym kiczem" – tanimi imitacjami głębszych idei, odcinającymi

nas od inteligentnego zaangażowania i osłabiającymi naszą kulturę. Kicz mówi nam, jak myśleć i jak się czuć. Handlarze kiczem odnoszą korzyści, podczas gdy konsumenci kiczu tracą. Wszyscy kończymy w spłyconym świecie.

W części drugiej, zatytułowanej *Problem wnioskowania*, twierdzę, że jedynym rodzajem wnioskowania, innymi słowy „myśleniem", którym musiałaby się posługiwać AI mająca odpowiadać poziomowi ludzkiemu (lub czemukolwiek, co się do tego zbliża), jest taki rodzaj, co do którego nie mamy pojęcia, jak go zaprogramować czy skonstruować. Problem wnioskowania dotyka samego serca debaty o AI, ponieważ dotyczy bezpośrednio inteligencji, zarówno ludzi, jak i maszyn. Nasza wiedza o różnych typach wnioskowania sięga Arystotelesa i innych starożytnych Greków i była rozwijana w dziedzinach logiki i matematyki. Wnioskowanie jest obecnie opisywane przy użyciu formalnych, symbolicznych systemów, takich jak programy komputerowe, więc analizując kwestię wnioskowania, można uzyskać bardzo wyraźny obraz projektu inżynierii inteligencji. Istnieją trzy typy wnioskowania. Klasyczne podejście do AI skupiało się na jednym typie – dedukcji, nowoczesne ujęcie AI zajmuje się innym – indukcją. Kluczowy dla ogólnej inteligencji jest trzeci typ – abdukcja, nad którym, niespodzianka, nikt nie pracuje – wcale[2]. Wreszcie, ponieważ każdy typ wnioskowania jest odrębny – co oznacza, że jeden typ nie może być zredukowany do innego – wiemy, że niepowodzenie w budowaniu systemów AI przy użyciu typu wnioskowania wspierającego ogólną inteligencję będzie skutkowało brakiem postępów w kierunku sztucznej ogólnej inteligencji (ang. *artificial general intelligence* – AGI).

W części trzeciej, zatytułowanej *Przyszłość mitu*, twierdzę, że mit ma bardzo niekorzystne konsekwencje, jeśli traktuje się go poważnie, ponieważ podważa naukę. W szczególności osłabia kulturę inteligencji ludzkiej i wynalazczości,

[2] Nie zamierzam sugerować, że naukowcy nie zmagali się z abdukcją w AI – zmagali się. W latach osiemdziesiątych i dziewięćdziesiątych XX wieku pracowano nad logicznymi podejściami do abdukcji, zwanymi „programowaniem opartym na logice abdukcyjnej". Jednak systemy te łączyła z abdukcją „tylko nazwa", ponieważ polegały na dedukcji, a nie na prawdziwej abdukcji. Systemy te nie odnosiły sukcesów i szybko zostały porzucone, gdy w pracach nad AI nadeszła era Internetu. Ostatnio, od około 2010 roku do dzisiaj, przyjęto różne probabilistyczne (w szczególności bayesowskie) podejścia jako możliwe ścieżki do autentycznego wnioskowania abdukcyjnego. Te systemy jednak również nie są w pełni abdukcyjne. Nie są one przebranymi podejściami dedukcyjnymi jak ich poprzednicy, ale są to przebrane podejścia indukcyjne lub probabilistyczne. Abdukcja tylko z nazwy nie jest tym, co mam na myśli, mówiąc o abdukcji, a systemy używające tej nazwy, ale nierozwiązujące problemu, nie pomogą nam poczynić postępów w AI. Wszystko to wyjaśnię na kolejnych stronach.

niezbędną dla osiągania istotnych przełomów, których będziemy potrzebować, aby zrozumieć naszą własną przyszłość. Nauka o danych (zastosowanie AI w kontekście „wielkich zbiorów danych" – ang. *big data*) jest w najlepszym wypadku protezą ludzkiej pomysłowości, która, używana prawidłowo, może pomóc nam poradzić sobie ze współczesnym „zalewem danych". Jeśli jednak jest wykorzystywana jako zamiennik dla indywidualnej inteligencji, ma tendencję do pochłaniania inwestycji bez przynoszenia oczekiwanych rezultatów. Wyjaśniam, w szczególności, jak wśród innych współczesnych przedsięwzięć naukowych mit negatywnie wpłynął na badania w neurobiologii. Cena, którą płacimy za mit, jest zbyt wysoka. Ponieważ nie mamy dobrego naukowego powodu, by wierzyć, że mit jest prawdziwy, a wszystko wskazuje na to, by go odrzucić na rzecz naszego własnego rozkwitu, musimy radykalnie przemyśleć dyskusję na temat AI.

Część I

Uproszczony świat

Rozdział 1

Błędne rozumienie inteligencji

Historia sztucznej inteligencji zaczyna się od idei człowieka, który posiadał ogromną inteligencję ludzką – pioniera komputerowego Alana Turinga. W 1950 roku Turing opublikował prowokacyjny artykuł *Maszyna licząca a inteligencja*[3], w którym poruszył możliwość istnienia inteligentnych maszyn. Artykuł był odważny, ponieważ ukazał się w czasach, kiedy komputery stanowiły nowość i, jak na dzisiejsze standardy, były mało imponujące. Wolne, ciężkie jednostki sprzętowe przyspieszały obliczenia naukowe, takie jak łamanie kodów. Po wielu przygotowaniach można było wprowadzić do nich równania fizyczne oraz warunki początkowe, a one obliczałyby promień wybuchu bomby nuklearnej. IBM szybko dostrzegł ich potencjał, jeśli chodzi o zastępowanie pracujących w przedsiębiorstwach ludzi zajmujących się obliczeniami, jak choćby aktualizowanie arkuszy kalkulacyjnych. Jednak postrzeganie komputerów jako „myślących" wymagało niemałej wyobraźni.

Propozycja Turinga opierała się na popularnej zabawie znanej jako „gra w udawanie"[4]. W oryginalnej grze mężczyzna i kobieta są ukryci przed wzrokiem innych. Trzecia osoba, przesłuchujący, zadaje pytania tylko jednej z osób na raz i, analizując odpowiedzi, próbuje określić, kto jest mężczyzną, a kto kobietą. Zabawa polega na tym, że mężczyzna musi próbować wprowadzić w błąd osobę przesłuchującą, podczas gdy kobieta stara się mu pomóc – to sprawia, że odpowiedzi płynące z obu stron stają się podejrzane. Turing zastąpił mężczyznę i kobietę komputerem i człowiekiem. Tak wyglądała geneza tego, co obecnie nazywamy testem Turinga, w którym komputer i człowiek otrzymują pytania w formie pisemnej od ludzkiego sędziego, a jeśli sędzia nie jest w stanie dokładnie określić, który z nich jest komputerem, komputer wygrywa.

[3] A.M. Turing, *Maszyna licząca a inteligencja*, tłum. M. Szczubiałka, w: *Filozofia umysłu*, wyb. i wstępem opatrzył B. Chwedeńczuk, Fundacja ALETHEIA – Wydawnictwo SPACJA, Warszawa 1995, s. 271–300.

[4] Tamże, s. 271.

Turing argumentował, że w takim przypadku nie mamy dobrego powodu, by uznawać maszynę za nieinteligentną, niezależnie od tego, że nie jest człowiekiem. W ten sposób pytanie o to, czy maszyna ma inteligencję, zastępuje pytanie o to, czy może naprawdę myśleć.

Test Turinga jest w rzeczywistości bardzo trudny – żaden komputer nigdy go nie przeszedł. Oczywiście w 1950 roku Turing nie wiedział, że do dzisiaj taki będzie wynik testu; jednak zastępując uporczywe filozoficzne pytania o „świadomość" i „myślenie" testem, który dotyczy widocznych wyników, zachęcił do postrzegania AI jako uprawnionej dziedziny nauki z jasno określonym celem. Gdy w latach pięćdziesiątych XX wieku zaczęły się badania nad AI, wielu pionierów i zwolenników idei AI zgodziło się z Turingiem, że w przypadku dowolnego komputera, który prowadziłby długotrwałą i przekonującą rozmowę z człowiekiem, należałoby, jak większość z nas by przyznała, stwierdzić, że wykonuje on coś, co wymaga myślenia (cokolwiek by to miało znaczyć).

Rozróżnienie Turinga między intuicją a pomysłowością

Turing zdobył uznanie jako matematyk na długo przed tym, nim zaczął pisać o AI. W 1936 roku opublikował krótki artykuł matematyczny dotyczący dokładnego znaczenia terminu „komputer", który w tamtym czasie odnosił się do osoby wykonującej ustaloną kolejność działań w celu uzyskania określonego wyniku (takiego jak wykonanie obliczenia)[5]. W tym artykule ideę ludzkiego komputera zastąpił ideą maszyny wykonującej tę samą pracę.

W artykule tym mamy do czynienia z matematyką na zaawansowanym poziomie. Jednak w swoich rozważaniach na temat maszyn Turing nie odnosił się do ludzkiego myślenia ani umysłu. Twierdził, że maszyny mogą działać automatycznie, a problemy, które rozwiązują, nie wymagają żadnej „zewnętrznej" pomocy ani inteligencji. Tę zewnętrzną inteligencję – czynnik ludzki – matematycy czasami nazywają „intuicją"[6].

---

[5] A.M. Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, „Proceedings of the London Mathematical Society" 1937, Vol. 2–42, issue 1, s. 230–265.

[6] Ang. słowo *insight* będzie w dalszej części książki tłumaczone jako „intuicja", ale też „zdolność do zrozumienia", „zrozumienie", „wgląd" itp. jako odnoszące się do zdolności do uchwycenia, zrozumienia jakiejś prawdy, jakiegoś zagadnienia, problemu itp. (przyp. tłum.).

Praca Turinga z 1936 roku dotycząca maszyn obliczeniowych pomogła zapoczątkować informatykę jako dyscyplinę i stanowiła ważny wkład w logikę matematyczną. Mimo to Turing najwyraźniej uważał, że jego wczesna definicja pomija coś istotnego. W rzeczy samej, ta sama idea umysłu lub ludzkich zdolności wspomagających rozwiązywanie problemów pojawiła się dwa lata później w jego pracy doktorskiej, będącej sprytną, ale ostatecznie nieudaną próbą obejścia wyniku otrzymanego przez austriackiego logika matematycznego Kurta Gödla (więcej o tym później). Praca Turinga zawiera następujący ciekawy fragment dotyczący intuicji, w którym porównuje ją z inną zdolnością mentalną, określoną przez niego jako „pomysłowość":

> Rozumowanie matematyczne można, upraszczając nieco sprawę, uznać za połączenie dwóch zdolności, które możemy nazwać intuicją i pomysłowością. Działalność intuicji polega na podejmowaniu spontanicznych osądów, które nie są wynikiem świadomego toku rozumowania. Te osądy są często, ale nie zawsze, poprawne (pomijając kwestię tego, co oznacza „poprawność"). Często istnieje możliwość znalezienia różnych sposobów weryfikacji poprawności intuicyjnego osądu. Można na przykład stwierdzić, że każda dodatnia liczba całkowita posiada jednoznaczny rozkład na liczby pierwsze; szczegółowe rozumowanie matematyczne prowadzi do tego samego wniosku. Będzie to również obejmowało intuicyjne osądy, jednak będą one mniej podatne na krytykę niż pierwotny osąd dotyczący rozkładu na czynniki pierwsze. Nie zamierzam jednak próbować wyjaśniać tej idei „intuicji" w sposób bardziej szczegółowy.

Turing następnie przechodzi do wyjaśnienia pomysłowości: „Wykorzystanie pomysłowości w matematyce polega na wspieraniu intuicji poprzez odpowiednie uporządkowanie twierdzeń, a czasem także za pomocą figur geometrycznych lub rysunków. Zakłada się, że jeśli te elementy zostaną naprawdę dobrze zorganizowane, poprawność wymaganych kroków intuicyjnych nie może budzić poważnych wątpliwości"[7].

Chociaż jego język jest skierowany do specjalistów, Turing wskazuje na coś oczywistego, mianowicie że matematycy zazwyczaj wybierają problemy lub „dostrzegają" interesujące zadania do rozwiązania, korzystając z jakiejś zdolności, w której, jak się wydaje, nie mamy do czynienia z poszczególnymi krokami – a zatem nie jest ona w oczywisty sposób podatna na programowanie komputerowe.

[7] A.M. Turing, *Systems of Logic Based on Ordinals*, rozprawa doktorska, Princeton University 1938, s. 57.

Idea Gödla

Gödel także myślał o inteligencji mechanicznej. Podobnie jak Turing miał obsesję na punkcie różnicy między pomysłowością (mechaniką) a intuicją (umysłem). Jego rozróżnienie było zasadniczo takie samo jak Turinga, lecz wyrażone innym językiem: dowód versus prawda (lub „teoria dowodów" versus „teoria modeli" w terminologii matematycznej). Gödel zastanawiał się, czy pojęcia dowodu i prawdy oznaczają ostatecznie to samo. Jeśli tak, matematykę, a nawet samą naukę można by rozumieć wyłącznie w sposób mechaniczny. Ludzkie myślenie w tym ujęciu również mogłoby być mechaniczne. Pojęcie AI, choć termin jeszcze nie powstał, krążyło wokół tego pytania. Czy cechująca umysł intuicja, jej zdolność do uchwycenia prawdy i znaczenia, może być zredukowana do tego, co robią maszyny, do obliczeń?

To pytanie stawiał sobie Gödel. W trakcie próby odpowiedzi na nie natrafił na pewną przeszkodę, co szybko przyczyniło się do zyskania przez niego światowej sławy. W 1931 roku Gödel opublikował dwa twierdzenia z logiki matematycznej, znane jako twierdzenia o niezupełności. Wykazał w nich istnienie wrodzonych ograniczeń wszystkich formalnych systemów matematycznych. To było genialne spostrzeżenie. Gödel jednoznacznie pokazał, że matematyka – **cała** matematyka, przy pewnych prostych założeniach – nie jest, w ścisłym ujęciu, mechaniczna ani „formalizowalna". Mówiąc bardziej precyzyjnie, Gödel udowodnił, że w każdym formalnym (matematycznym lub obliczeniowym) systemie muszą istnieć pewne zdania, które są Prawdziwe – „P" pisane wielką literą, a jednak nie mogą zostać udowodnione w tym systemie za pomocą jakichkolwiek z jego reguł. Prawdziwe zdanie może być rozpoznane przez ludzki umysł, ale (co można udowodnić) jego prawdziwości nie da się udowodnić w systemie, w którym zostało sformułowane.

Jak Gödel doszedł do tego wniosku? Szczegóły są skomplikowane i techniczne, ale podstawowa idea Gödla polega na tym, że możemy traktować system matematyczny, wystarczająco złożony, by wykonywać operacje dodawania, jako system znaczeń, niemal jak naturalny język, taki jak angielski czy niemiecki – to samo dotyczy wszystkich bardziej złożonych systemów. Traktując go w ten sposób, umożliwiamy systemowi mówienie o sobie. Może on na przykład stwierdzić, że ma pewne ograniczenia. To właśnie dostrzegł Gödel.

Systemy formalne, takie jak te w matematyce, pozwalają na precyzyjne wyrażenie prawdy i fałszu. Zazwyczaj ustalamy prawdę, korzystając z narzędzi dowodowych – używamy reguł, aby coś udowodnić, dzięki czemu wiemy, że jakieś twierdzenie jest z całą pewnością prawdziwe. Ale czy istnieją prawdziwe stwierdzenia, których nie można udowodnić? Czy umysł może znać rzeczy, których system znać nie może? W arytmetyce wyrażamy prawdy, zapisując równania takie jak: „2 + 2 = 4". Zwykłe równania są prawdziwymi twierdzeniami w systemie arytmetyki i mogą zostać udowodnione zgodnie z regułami arytmetycznymi. W tym wypadku to, co można udowodnić, jest równoznaczne z prawdą. Matematycy przed Gödelem uważali, że własność tę ma cała matematyka. To sugerowało, że maszyny po prostu stosując reguły w sposób poprawny, mogłyby wygenerować wszystkie prawdy w różnych systemach matematycznych. To piękna idea, ale nieprawdziwa.

Gödel odkrył rzadką, ale potężną właściwość samoodniesienia. Bez naruszania zasad systemów matematycznych można skonstruować matematyczne wersje wyrażeń odwołujących się do samych siebie, takich jak: „To zdanie nie jest dowodliwe w tym systemie". Jednak tak zwane „zdania Gödla", gdy odnoszą się do samych siebie, wprowadzają do matematyki sprzeczność. Jeśli są prawdziwe, to są niedowodliwe. Jeśli są fałszywe, to ponieważ twierdzą, że są niedowodliwe, w rzeczywistości są prawdziwe. Prawda oznacza fałsz, a fałsz oznacza prawdę, co prowadzi do sprzeczności.

Wracając do pojęcia intuicji, my, ludzie, możemy dostrzec, że zdanie Gödla jest tak naprawdę prawdziwe, ale z powodu konsekwencji twierdzenia Gödla wiemy, że zasady systemu nie mogą go dowieść – system w zasadzie jest ślepy na coś, co nie jest objęte jego zasadami[8]. Prawda i dowodliwość zaczynają się od siebie oddalać. Może to również dotyczy działania umysłu i maszyn. Czysto formalny system ma swoje ograniczenia. Nie zdoła dowieść we własnym języku czegoś, co jest prawdziwe. Innymi słowy, możemy dostrzegać coś, czego komputer dostrzec nie potrafi[9].

[8] Gödel wykazał również, że dodanie nowych reguł mogłoby w niektórych systemach załatać tę niedoskonałość, ale nowy system, z dodatkowymi regułami, miałby z kolei inne ograniczenia, i tak w nieskończoność. To było dokładnie to, na czym Turing skoncentrował się w swojej późniejszej pracy nad systemami formalnymi i zupełnością.

[9] W sprawie oryginalnych wyników dotyczących twierdzeń o niezupełności zob. pracę K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, „Monatshefte

Ustalenia Gödla zadały potężny cios popularnej wówczas idei, że cała matematyka mogłaby być przełożona na operacje oparte na regułach, generując matematyczne prawdy jedna po drugiej. Duch czasów sprzyjał formalizmowi – nie było mowy o umysłach, duchach, duszach itp. Ruch formalistyczny w matematyce był wyrazem szerszego zwrotu intelektualistów w kierunku materializmu naukowego, a w szczególności pozytywizmu logicznego – ruchu poświęconego zwalczaniu tradycyjnej metafizyki, takiej jak platonizm z jego abstrakcyjnymi Formami, które nie mogły być postrzegane zmysłami, oraz tradycyjnych idei religijnych, takich jak idea istnienia Boga. Świat zwracał się w stronę koncepcji precyzyjnych maszyn. Nikt jednak nie podjął sprawy formalizmu tak energicznie jak niemiecki matematyk David Hilbert.

Wyzwanie Hilberta

Na początku XX wieku (jeszcze przed ogłoszeniem twierdzeń Gödla) Hilbert postawił wyzwanie światu matematycznemu: pokażcie, że cała matematyka opiera się na solidnych fundamentach. Obawa Hilberta była zrozumiała. Jeśli czysto formalne zasady matematyki nie mogą dowieść wszystkich prawd, to teoretycznie możliwe jest, że matematyka ukrywa sprzeczności i nonsens. Sprzeczność ukryta gdzieś w matematyce psuje wszystko, ponieważ ze sprzeczności wynika logicznie cokolwiek. Wówczas formalizm staje się bezużyteczny. Hilbert wyraził marzenie wszystkich formalistów, aby w końcu udowodnić, że matematyka jest zamkniętym systemem rządzonym wyłącznie przez reguły. Prawda to tylko „dowód". Zyskujemy wiedzę, po prostu śledząc „kod" dowodu i potwierdzając, że żadne reguły nie zostały naruszone. Większe marzenie, skrycie zamaskowane, było w rzeczywistości światopoglądem – obrazem wszechświata jako mechanizmu. Pomysł sztucznej inteligencji zaczął kształtować się jako idea, filozoficzne stanowisko, które również mogłoby być udowodnione. Formalizm traktował inteligencję jako proces oparty na regułach; jako działanie maszyny.

Hilbert postawił swoje wyzwanie podczas Drugiego Międzynarodowego Kongresu Matematyków w Paryżu w 1900 roku. Świat intelektualny słuchał. Jego wyzwanie miało trzy główne części: udowodnić, że matematyka jest zupełna;

für Mathematik Physik" 1931, 38, s. 173–198. Tłumaczenie na język angielski: *On formally undecidable propositions of Principia mathematica and related systems I*, w: *K. Gödel. Collected Works,* Vol. 1: 1929–1936, eds. S. Feferman *et al.*, Oxford University Press, Oxford 1986, s. 145–195.

udowodnić, że matematyka jest spójna; oraz udowodnić, że matematyka jest rozstrzygalna.

Gödel zadał śmiertelny cios pierwszym i drugim częściom wyzwania Hilberta, publikując w roku 1931 swoje twierdzenia o niezupełności. Pytanie o rozstrzygalność pozostało bez odpowiedzi. System jest rozstrzygalny, jeśli istnieje określona procedura (dowód lub sekwencja deterministycznych, oczywistych kroków), pozwalających ustalić, czy jakiekolwiek twierdzenie skonstruowane za pomocą zasad systemu jest prawdziwe czy fałszywe. Stwierdzenie 2 + 2 = 4 musi być prawdziwe, a 2 + 2 = 5 musi być fałszywe. Tak samo musi być dla wszystkich twierdzeń, które można poprawnie sformułować za pomocą symboli i reguł systemu. Ponieważ arytmetyka była uznawana za fundament matematyki, udowodnienie, że matematyka jest rozstrzygalna, oznaczało udowodnienie tego wyniku dla arytmetyki i jej rozszerzeń. Oznaczałoby to stwierdzenie, że matematycy, grając w „grę" z regułami i symbolami (idea formalistyczna), faktycznie grają w słuszną grę, która nigdy nie prowadzi do sprzeczności ani absurdu.

Turing był zafascynowany osiągnięciem Gödla, który wykazał nie moc systemów formalnych, lecz ich ograniczenia. Zajął się pracą nad pozostałą częścią wyzwania Hilberta i zaczął poważnie myśleć o tym, czy może istnieć procedura rozstrzygania dla systemów formalnych. W 1936 roku, w swoim artykule *Computable Numbers* udowodnił, że z pewnością nie istnieje. Turing dostrzegł, że wskazany przez Gödla problem samoodniesienia dotyczy także pytań z obszaru procedur rozstrzygania, a w konsekwencji programów komputerowych. W szczególności uświadomił sobie, że muszą istnieć (rzeczywiste) liczby, których żadna określona metoda nie byłaby w stanie „obliczyć", pisząc ich rozwinięcie dziesiętne, cyfra po cyfrze. Odniósł się do wyniku dziewiętnastowiecznego matematyka Georga Cantora, który udowodnił, że liczby rzeczywiste (te z rozwinięciem dziesiętnym) są liczniejsze niż liczby całkowite, mimo że zarówno liczby rzeczywiste, jak i całkowite są nieskończone. Turing być może stał na ramionach gigantów, jednak ostatecznie w *Computable Numbers* udowodnił, że dla problemu rozstrzygalności również uzyskujemy wynik negatywny. Rezultat ten oznaczał ograniczenia: nie jest możliwe stworzenie uniwersalnej procedury rozstrzygania o prawdziwości. Innymi słowy, reguły – nawet w matematyce – nie są wystarczające. Hilbert się mylił[10].

[10] Używam tutaj wymiennie terminów „formalny", „matematyczny" i „obliczeniowy". Terminologia ta nie jest niewłaściwa, chociaż technicznie wszystkie systemy matematyczne lub obliczeniowe znane są jako systemy formalne. Mam nadzieję, że nie wprowadza zamieszania, ale w każdym

Konsekwencje dla idei AI

Co istotne dla idei AI w tym kontekście, to fakt, że Turing obalił tezę, iż matematyka może być rozstrzygalna, wynajdując maszynę – maszynę deterministyczną – która nie wymagała żadnego wglądu ani inteligencji do rozwiązywania problemów. Dziś odnosimy się do jego abstrakcyjnej koncepcji maszyny jako do maszyny Turinga. Właśnie teraz piszę na takiej maszynie. Maszyny Turinga to komputery. To jedna z wielkich ironii historii intelektualnej, że teoretyczne ramy obliczeń zostały ustanowione jako myśl poboczna, środek do osiągnięcia innego celu. Pracując nad obaleniem tezy, że sama matematyka jest rozstrzygalna, Turing najpierw stworzył coś precyzyjnego i mechanicznego – komputer.

W swojej pracy doktorskiej z 1938 roku Turing miał nadzieję, że systemy formalne mogą zostać rozszerzone poprzez dodanie nowych reguł (później zbiorów reguł i zbiorów reguł), które mogłyby poradzić sobie z „problemem Gödla". Odkrył jednak, że nowy, potężniejszy system będzie miał nowy, bardziej skomplikowany „problem Gödla". Nie było sposobu na obejście twierdzenia Gödla o niezupełności. W skomplikowanym omówieniu przez Turinga tematu systemów formalnych tkwi jednak dziwna sugestia, istotna dla możliwości AI. Być może zdolność intuicji nie może być zredukowana do algorytmu, do reguł systemu?

W swojej rozprawie doktorskiej z 1938 roku Turing starał się znaleźć sposób na wyjście z ograniczeń będących konsekwencją odkryć Gödla, ale odkrył, że to niemożliwe. Zmienił więc kierunek swoich badań, starając się, jak sam to ujął, „znacząco zmniejszyć" potrzebę ludzkiej intuicji w procesie obliczeń. W swojej rozprawie rozważał możliwość pojawienia się pomysłowości w wyniku tworzenia coraz bardziej skomplikowanych systemów reguł. (Jak się okazało, pomysłowość może stać się uniwersalna – istnieją maszyny, które mogą przyłączać jako dane wejściowe inne maszyny, a tym samym uruchamiać wszystkie maszyny, jakie zbudowano. To odkrycie – technicznie mówiąc, dotyczące uniwersalnej maszyny Turinga, a nie prostej maszyny Turinga – miało stać się fundamentem komputera cyfrowego). Jednak w swojej formalnej pracy nad obliczeniami Turing, być może nieświadomie, ujawnił pewne istotne kwestie.

---

razie terminy „matematyczny" i „obliczeniowy" odnoszą się do systemów formalnych, które posiadają dobrze zdefiniowany słownik symboli oraz zasady manipulacji tymi symbolami. Mam więc na myśli zarówno arytmetykę, jak i języki komputerowe, wypowiadając się w sposób ogólny i dzięki temu dostosowany do celów niniejszego omówienia.

Przez odróżnienie intuicji od operacji w czysto formalnym systemie, takim jak komputer, Turing zasugerował, że mogą istnieć różnice między programami komputerowymi wykonującymi obliczenia a matematykami.

Dlatego też interesujący był zwrot, który Turing wykonał w stosunku do swojej wczesnej pracy z lat trzydziestych XX wieku w stronę szerszej spekulacji na temat możliwości inteligentnych komputerów, w artykule *Maszyny liczące a inteligencja*, opublikowanym nieco ponad dekadę później.

Do 1950 roku dyskusja na temat intuicji zniknęła z pism Turinga dotyczących implikacji wyników Gödla. Jego zainteresowania przesunęły się w stronę możliwości, że komputery mogą stać się „maszynami intuicyjnymi". W istocie stwierdził, że wyniki Gödla nie odnoszą się do pytania o AI: jeśli my, ludzie, jesteśmy wysoce zaawansowanymi komputerami, to wyniki Gödla oznaczają jedynie, że są pewne zdania, których nie możemy zrozumieć ani uznać za prawdziwe, podobnie jak w przypadku mniej skomplikowanych komputerów. Te zdania mogą być fantastycznie skomplikowane i interesujące lub być może banalne, ale jednocześnie przytłaczająco złożone. Wnioski Gödla pozostawiły otwartą kwestię, czy umysły są jedynie bardzo skomplikowanymi maszynami z równie skomplikowanymi ograniczeniami.

Innymi słowy, pojęcie intuicji stało się częścią idei Turinga o maszynach i ich możliwościach. Twierdzenia Gödla nie mogły (przynajmniej według Turinga) rozstrzygać o tym, czy umysły są maszynami, czy nie. Z jednej strony, twierdzenie o niezupełności mówi, że niektóre zdania mogą być postrzegane jako prawdziwe dzięki intuicji, ale nie mogą być dowiedzione przez komputer, który korzysta z pomysłowości. Z drugiej strony, bardziej zaawansowany komputer może użyć więcej aksjomatów (lub więcej fragmentów odpowiedniego kodu) i dowieść wyniku, co pokazuje, że intuicja nie stoi w sprzeczności z obliczeniami dokonywanymi w ramach danego problemu. Mamy do czynienia z wyścigiem zbrojeń: coraz potężniejsza pomysłowość zastępuje intuicję w coraz bardziej skomplikowanych problemach. Nikt nie jest w stanie powiedzieć, kto wygrywa ten wyścig, więc nikt nie może przedstawić argumentu – korzystając z twierdzenia o niezupełności – na rzecz wrodzonych różnic między intuicją (umysłem) a pomysłowością (maszyną). Jednak, jak na pewno wiedział Turing, jeśli to byłoby prawdą, to sztuczna inteligencja także byłaby przynajmniej możliwa.

W ten sposób, w latach 1938–1950, Turing zmienił swoje zdanie na temat pomysłowości i intuicji. W 1938 roku intuicja była tajemniczą „mocą

selekcji", która pomagała matematykom decydować, z jakimi systemami pracować i jakie problemy rozwiązywać. Intuicja nie była rzeczą w komputerze, lecz czymś, co decydowało o możliwościach komputera. W 1938 roku Turing uważał, że intuicja nie jest częścią żadnego systemu, co sugerowało nie tylko, że umysły i maszyny są zasadniczo różne, ale także, że AI jako ludzki sposób myślenia była niemal niemożliwa.

Jednak do 1950 roku odwrócił swoje stanowisko. Swoim testem Turinga postawił wyzwanie sceptykom oraz wyraził swego rodzaju obronę idei o możliwości intuicji w maszynach, pytając w istocie: „Dlaczego nie?". To był radykalny zwrot. Wydawało się, że rodzi się nowy pogląd na inteligencję.

Skąd ta zmiana? W latach 1938–1950 Turingowi przydarzyło się coś, co przyszło spoza świata ścisłej matematyki, logiki i systemów formalnych. W rzeczywistości zdarzyło się to całej Wielkiej Brytanii, a nawet większości świata. Była to druga wojna światowa.

Rozdział 2

Turing w Bletchley

Gra w szachy fascynowała Turinga – podobnie jak jego kolegę z czasów wojny, matematyka Irvina Johna „Jacka" Gooda. Ci dwaj panowie grali ze sobą (Good zazwyczaj wygrywał) i opracowywali procedury decyzyjne oraz zasady dla zwycięskich ruchów. Gra w szachy polega na stosowaniu zasad gry (korzystając z pomysłowości), a także wydaje się wymagać wglądu (intuicji) w to, które zasady wybrać w poszczególnych sytuacjach na planszy. Aby wygrać w szachy, nie wystarczy stosować zasady; trzeba przede wszystkim wiedzieć, które reguły wybrać.

Turing postrzegał szachy jako praktyczny (i niewątpliwie dający rozrywkę) sposób na myślenie o maszynach oraz o możliwości obdarzenia ich intuicją. Po drugiej stronie Atlantyku założyciel nowoczesnej teorii informacji, kolega i przyjaciel Turinga, Claude Shannon z Bell Labs, również myślał o szachach. Później zbudował jeden z pierwszych komputerów grających w szachy, będący rozwinięciem jego wcześniejszej pracy nad protokomputerem, zwanym „analizatorem różnicowym", który potrafił przekształcać niektóre problemy z rachunku różniczkowego w procedury mechaniczne[11].

Uproszczone ujęcie istot inteligentnych

Szachy fascynowały Turinga i jego współpracowników częściowo dlatego, że wydawało się, iż komputer można zaprogramować do gry w nie bez potrzeby, aby ludzki programista przewidział z góry wszystkie możliwe ruchy.

[11] Wczesna praca Turinga, Gooda i Shannona nad komputerowym graniem w szachy wykorzystywała technikę znaną jako „minimax", która oceniała ruchy na podstawie minimalizacji strat dla gracza przy jednoczesnej maksymalizacji potencjalnych zysków. Technika ta miała ważne znaczenie w późniejszych wersjach szachów komputerowych i nadal stanowi punkt odniesienia dla projektowania znacznie bardziej zaawansowanych systemów komputerowych używanych dzisiaj do gry w szachy.

Ponieważ urządzenia obliczeniowe implementowały operatory logiczne, takie jak „jeśli–to", „lub" oraz „i", program (zestaw instrukcji) mógł być uruchamiany, a jego wyniki mogły się różnić w zależności od scenariuszy, jakie napotykał podczas wykonania instrukcji. Ta zdolność do zmiany kursu w zależności od tego, co komputer „widział", wydawała się Turingowi i jego współpracownikom symulować kluczowy aspekt ludzkiego myślenia[12].

Gracze w szachy – Turing, Good, Shannon i inni – myśleli również o innym problemie matematycznym, którego stawka była znacznie wyższa. W ramach pracy dla swoich rządów pomagali złamać tajne kody używane przez Niemcy do koordynowania ataków na statki handlowe i wojskowe przepływające przez Kanał La Manche i Atlantyk. Turing był zaangażowany w desperacką próbę pomocy w pokonaniu nazistowskich Niemiec podczas drugiej wojny światowej, i to jego pomysły dotyczące obliczeń przyczyniły się do zmiany jej biegu.

Bletchley Park

Bletchley Park, położony nieco na uboczu w małym miasteczku, z dala od bombardowań, z jakimi borykał się Londyn i obszary metropolitalne Wielkiej Brytanii, był miejscem badań służącym odkrywaniu lokalizacji niemieckich U-bootów – okrętów podwodnych, które niszczyły szlaki żeglugowe w Kanale La Manche. U-booty były poważnym problemem dla sił alianckich, ponieważ zatapiały tysiące statków i niszczyły ogromne ilości zaopatrzenia i sprzętu. Aby kontynuować działania wojenne, Wielka Brytania potrzebowała importu wynoszącego 30 milionów ton rocznie. U-booty w pewnym momencie zmniejszyły tę liczbę o 200 tysięcy ton miesięcznie, co stanowiło istotną, potencjalnie katastrofalną strategię Niemiec, z którą przez jakiś nie potrafiono sobie poradzić. W odpowiedzi na tę sytuację rząd brytyjski zebrał grupę utalentowanych kryptologów, szachistów i matematyków, aby zbadać, jak złamać komunikację U-bootów, o której wiadomo było, że jest zaszyfrowana. (Szyfr

[12] Pełne języki programowania, jakie znamy dzisiaj, takie jak C++ i Java, korzystają z podstawowych operacji, które wyłoniły się z wczesnych obliczeń, chociaż takie idee jak programowanie obiektowe i inne metody strukturyzacji oprogramowania pojawiły się później w informatyce. Mimo to podstawowe struktury sterujące we wszelkich kodach komputerowych pojawiły się już na początku, wraz z pierwszymi pełnymi maszynami elektronicznymi. Niewątpliwie do szybkiego sukcesu takich systemów, stosowanych we wczesnych problemach, jak szachy przyczynił się wgląd w to, jak strukturyzować i kontrolować maszyny za pomocą programów.

to zakamuflowana wiadomość. Zdeszyfrowanie wiadomości oznacza przywrócenie jej do czytelnego tekstu)[13].

Kody były generowane przez urządzenie przypominające maszynę do pisania, znane jako Enigma, które było w użyciu komercyjnym od lat dwudziestych XX wieku, ale Niemcy znacząco ulepszyli Enigmę w celu stosowania jej podczas wojny. Zmodyfikowane Enigmy były używane do wszystkich strategicznych komunikacji w niemieckich działaniach wojennych. Przykładowo, Luftwaffe korzystała z maszyny Enigma w przebiegu działań powietrznych, podobnie jak Kriegsmarine w operacjach morskich. Wiadomości zaszyfrowane zmodyfikowaną Enigmą powszechnie uważano za nie do odszyfrowania.

Rola Turinga w Bletchley i jego późniejsza popularność jako bohatera narodowego po wojnie to wielokrotnie opowiedziana historia. (Jego pracę w Bletchley oraz późniejsze działania na rzecz rozwoju komputerów zobrazował w 2014 roku film *Gra tajemnic*). Główne osiągnięcie Turinga, według czysto matematycznych standardów, było stosunkowo mało interesujące, ponieważ wykorzystywało starą ideę z logiki dedukcyjnej. Metoda, którą on i inni żartobliwie nazywali „Turingizmem", polegała na eliminowaniu dużych ilości możliwych rozwiązań kodów Enigmy poprzez znajdowanie kombinacji ze sprzecznościami. Kombinacje sprzeczne są niemożliwe; w systemie logicznym nie możemy mieć jednocześnie zarówno „A", jak i „nie-A", podobnie jak nie możemy być jednocześnie „w sklepie" i „w domu". Turingizm okazał się zwycięską ideą i odniósł ogromny sukces w Bletchley. „Geniuszom" skrytym w *think tanku* umożliwił przyspieszenie zadania odszyfrowania wiadomości Enigmy. Inni naukowcy opracowali różne strategie łamania kodów w Bletchley[14]. Pomysły były testowane na maszynie zwanej „Bombe" – jej żartobliwa nazwa pochodzi z wcześniejszej maszyny używanej w Polsce, znanej jako Bomba, i prawdopodobnie została zainspirowana krótkim dźwiękiem wydawanym po zakończeniu obliczenia. Można myśleć o Bombe jako o protokomputerze zdolnym do uruchamiania różnych programów.

Siły alianckie zaczęły mieć przewagę w wojnie w lub około 1943 roku, w dużej mierze dzięki nieustannemu wysiłkowi łamaczy kodów z Bletchley. Zespół odniósł wielki sukces, a jego członkowie stali się bohaterami wojennymi.

[13] Działo się to pod kierownictwem General Code and Cipher School [Główna Szkoła Kodów i Szyfrów] w Wielkiej Brytanii, znanej również jako „GC and CS".

[14] Na przykład Hugh Alexander zaangażowany w pracę w Bletchley był krajowym mistrzem szachowym.