

RAPORT POWSTAŁ NA ZLECENIE



ARTIFICIAL INTELLIGENCE

SumUp for Business 2022



Dr Anna Gumkowska, Warner Bros. Discovery

Sebastian Kondracki, Deviniti.com



1. Wprowadzenie	3	5.4 Speech recognition	40
2. Definicje AI	6	5.5 Robotyka	42
3. Krótka historia AI	9	6. Hiper-automatyzacja	44
4. Przegląd technik i algorytmów	14	7. Pułapki AI	46
4.1 Nadzorowane uczenie maszynowe	16	8. Casebook	50
4.2 Nienadzorowane uczenie maszynowe	18	8.1 Reklama internetowa	52
4.3 Uczenie maszynowe ze wzmocnieniem	20	8.2 Zdrowie	62
4.4 Głębokie uczenie	22	8.3 E-commerce	67
GAN – Generative Adversarial Network	25	8.4 Prawo	75
Transformery	26	8.5 Dobro społeczne	80
5. Domeny AI	29	9. Podsumowanie	83
5.1 Data science	31	10. O autorach	86
5.2 NLP	33		
5.3 Computer Vision	37		

S P I S T R E Ś C I

1.

WPROWADZENIE



Celem publikacji jest zebranie wiedzy dotyczącej AI i pokazanie jej praktycznych zastosowań w biznesie.

Według danych Microsoft 2/3 firm w Polsce już wykorzystuje AI¹. Nie zawsze zdajemy sobie nawet z tego sprawę, że w naszej organizacji już kilka zespołów na co dzień pracuje z AI. Nastąpiła cicha rewolucja. Początkowe ogromne oczekiwania i wyobrażenia związane z AI ulegały modyfikacjom. Na pewnych etapach inwestorzy tracili nawet nadzieję, że idee sztucznej inteligencji uda się wcielić w życie. Dzisiaj nie szukamy już jednej inteligencji, która porozmawia z nami o problemach egzystencjalnych przy kawie, za to mamy szereg konkretnych rozwiązań wymagających rozwiązań z zakresu AI, które towarzyszą nam w technologiach codziennego użytku.

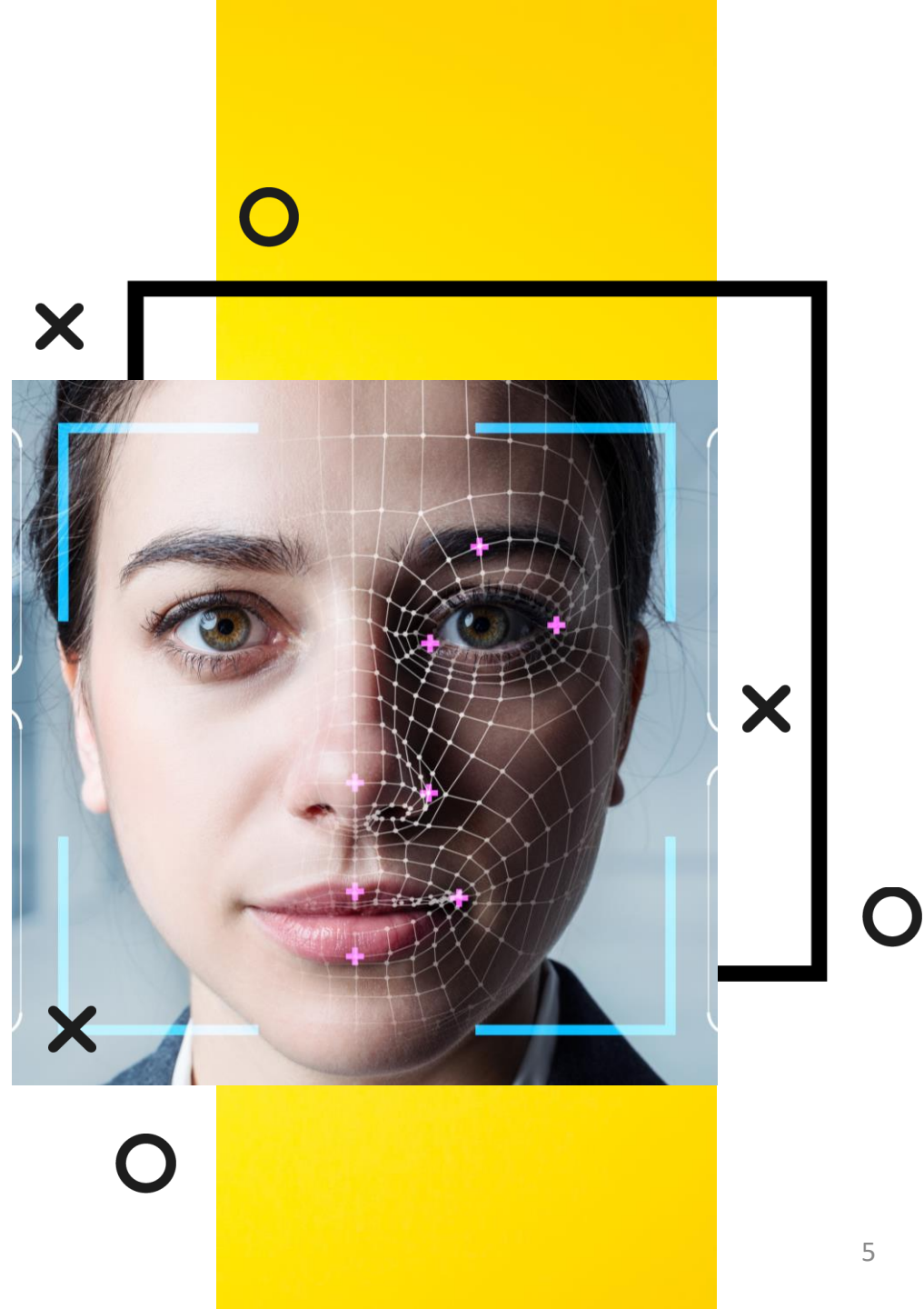
Sztuczna inteligencja stała się wszechobecna. Integruje się z wieloma dziedzinami ludzkiego życia: od stawiania diagnoz medycznych, przez marketing, social media, projekty społeczne, odkrywanie leków czy zastosowania w kancelariach prawnych i działach księgowości. AI przekształca naszą rzeczywistość – po cichu, ale systematycznie. To, że mamy w domu rozwiązania GPTs, czyli general-purpose technologies, jak automatyczny odkurzacz, mop czy żelazko rozpoznające materiał i samo dobierające odpowiednio temperaturę, nie dziwi już nikogo. Nieco więcej kontrowersji wzbudza zastosowanie facial recognition, automatyczne tworzenie kontentu czy AI zastosowany w sztuce, gdzie sam generuje obrazy dzięki GAN (Generative Adversarial Network).

¹ <https://przemyslprzyszlosci.gov.pl/microsoft-2-3-duzych-firm-w-polsce-juz-stosuje-ai/>

Ten dokument jest skierowany do osób, które chcą w skrócie zobaczyć najbardziej aktualny „duży obrazek”.

Dotykamy także zagadnień stanowczo obecnych w dyskusjach technologów, ale dopiero „pączkujących” w biznesie, takich jak **Transformery, LLM (ang. Large Language Models, m.in. GPT-3, Cohere, PaLM, Bloom)** i **hiper-automatyzacja**, które w ostatnim czasie całkowicie zrewolucjonizowały AI. Nie ma ambicji wchodzenia w szczegóły technologiczne ani nie wymaga wiedzy programistycznej. Ma za zadanie pomóc rozeznaczyć się w aktualnej sytuacji tak, by wzbudzić uważność, gdzie na co dzień mamy kontakt z AI. Ma także ułatwić zrozumienie, gdzie sami możemy AI zaimplementować z korzyścią dla naszego biznesu. Przede wszystkim jednak ma sprawić, że tematyka AI zostanie odczarowana: oczywiście, to są zaawansowane technologie, trudno wyobrażalne moce obliczeniowe i klastry danych. Natomiast nie jest to czarna magia. Na co dzień spotykamy się ze sztuczną inteligencją, kiedy tylko odpalimy smartfona.

Inspiracją do powstania tego raportu był kurs **Artificial Intelligence Programme for Business na Saïd Business School of Oxford**, a także doświadczenia w firmie Deviniti, która realizuje projekty w duchu strategii hiper-automatyzacji zwłaszcza w obszarze przetwarzania i generowania języka naturalnego ze słynnymi LLM (GPT-3, GPT-NeoX) na czele.



2.

DEFINICJE AI

Zanim dojdziemy do definicji (zwykle definicje są problematyczne, niejednoznaczne i nie są w stanie uchwycić wszystkiego), anegdota z życia znajomego z branży mediów. Jarek, bo to o nim tu mowa, regularnie testuje, czy dzwoniące do niego głosy z różnymi ofertami należą do ludzi czy do botów. Pierwszy etap testu zakłada postawienie pytania: „czy jest pani botem?”. Po ostatnich zmianach w oprogramowaniu odpowiedź brzmi: „nie, tylko tak dzwonię od rana po ludziach, to mój głos tak brzmi. Pewnie do pana też tak różne osoby dzwonią. Botem nie jestem, na dowód mogę powiedzieć, że mój rozmiar buta to 38”. Jarek nie daje jednak za wygraną: „a w jakim kolorze nosi Pani kapelusz?” Tutaj system się zawiesza. Bot nie ma przygotowanego scenariusza na tak skomplikowane pytania. Zapada długotrwała cisza. Jarek już wie, że dzwonił do niego automat oparty o algorytmy.

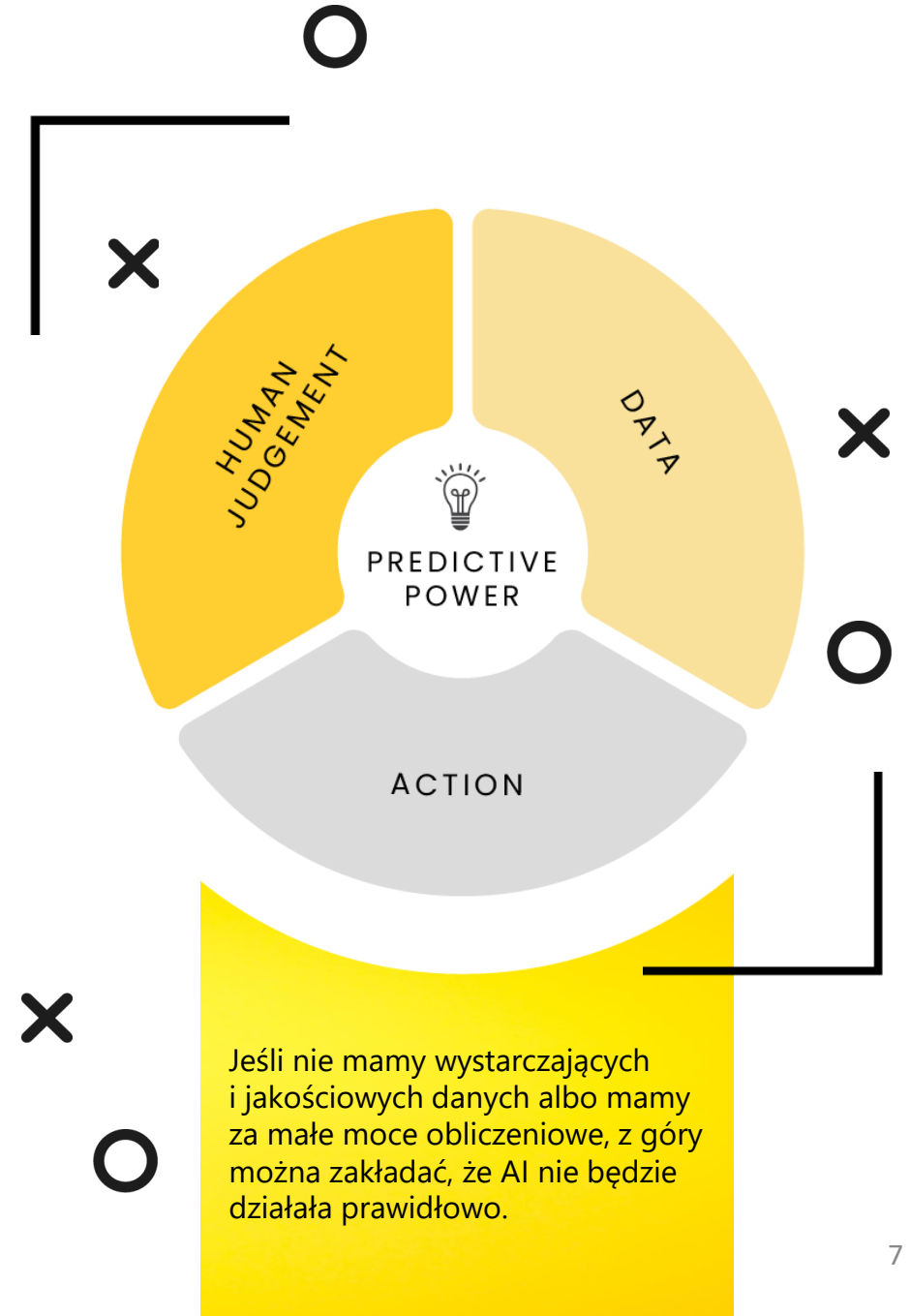
”

Wróćmy zatem do pytania, czym jest AI. Istotą sztucznej inteligencji jest zdolność maszyn do interpretowania danych i uczenia się na ich podstawie po to, by móc lepiej przygotować różnego typu predykcje. Warto podkreślić, że algorytmy uczą się w dużej mierze same: analizują zebrane dane, wyciągają wnioski, testują opcje i rekomendują te optymalne. Z jednej strony algorytmy są jak gotowy przepis napisany po to, by osiągnąć określone cele. Z drugiej strony same uczą się, rozwijają, ewoluują właśnie po to, by ten cel osiągnąć jak najlepiej. **Sztuczna inteligencja ma nam przede wszystkim pomóc w:**

- podejmowaniu decyzji,
- przygotowywaniu predykcji,
- automatyzacji tam, gdzie jest to możliwe, sensowne i ma uzasadnienie biznesowe.

Czego potrzebujemy, by działania AI były optymalne:

- Danych ustrukturyzowanych albo nieustrukturyzowanych – jakość danych, na których bazuje algorytm, jest ogromnie ważna. Obowiązuje tu prosta zasada: shit in, shit out.
- Big data i mocy obliczeniowej. Złożoność obliczeniowa często rośnie wykładniczo. Im więcej zmiennych masz w swojej analizie, tym więcej danych potrzebujesz, aby wyciągnąć prawidłowe wnioski.



Jeśli nie mamy wystarczających i jakościowych danych albo mamy za małe moce obliczeniowe, z góry można zakładać, że AI nie będzie działała prawidłowo.

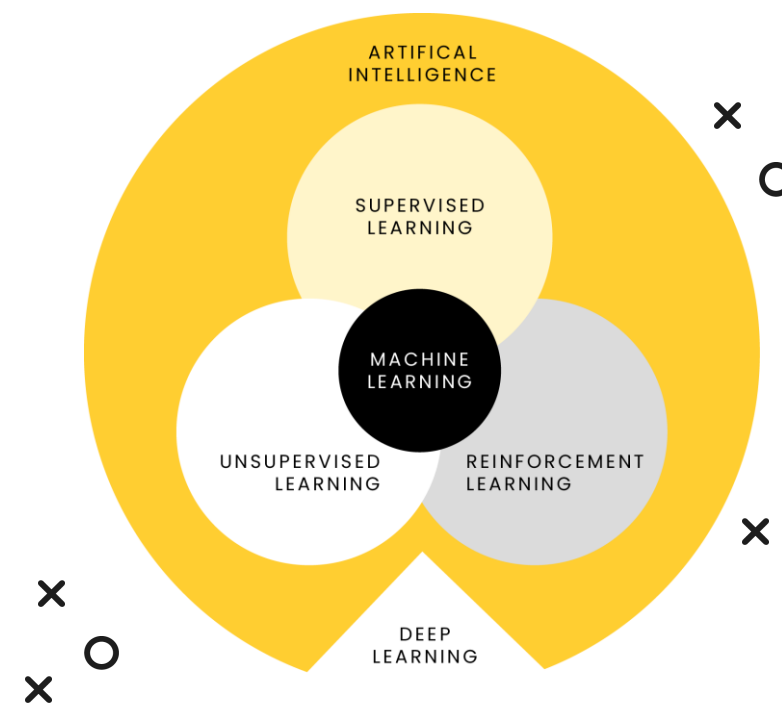
Czy da się rozwiązać każde zadanie czy problem dzięki AI?

Nie.

Istnieje grupa problemów nazywanych NP-trudnymi, których rozwiązanie w akceptowalnym czasie jest niemożliwe. Algorytm musi przeglądać wszystkie możliwe warianty, a ich liczba jest ogromna. To, że dziś AI nie potrafi sobie z niektórymi poradzić, nie oznacza, że za dziesięć lat będzie to wciąż niemożliwe. Gra w Go uważana była za jeden z najtrudniejszych problemów strategicznych i choć oficjalnie nie jest zaliczana do problemów NP-trudnych, to wielu ekspertów uważało, że AI długo sobie z nią nie poradzi. Jednak w 2016 roku program AlphaGo, stworzony przez OpenAI, pokonał jednego z najlepszych na świecie graczy w grze Go, co było ważnym krokiem w rozwoju AI i pokazało, że modele AI są w stanie pokonać nawet najlepszych ludzkich graczy w trudnych grach strategicznych.

Jeśli chodzi o definicję, przyjmuje się, że AI jest terminem najszerszym, które obejmuje takie kategorie, jak: machine learning (uczenie maszynowe), supervised learning (nadzorowane uczenie maszynowe), unsupervised learning (nienadzorowane uczenie maszynowe), reinforcement learning (uczenie maszynowe ze wzmocnieniem), a także deep learning (głębokie uczenie). Wzajemny układ elementów ilustruje wykres.

Wzajemny układ elementów ilustruje wykres poniżej.

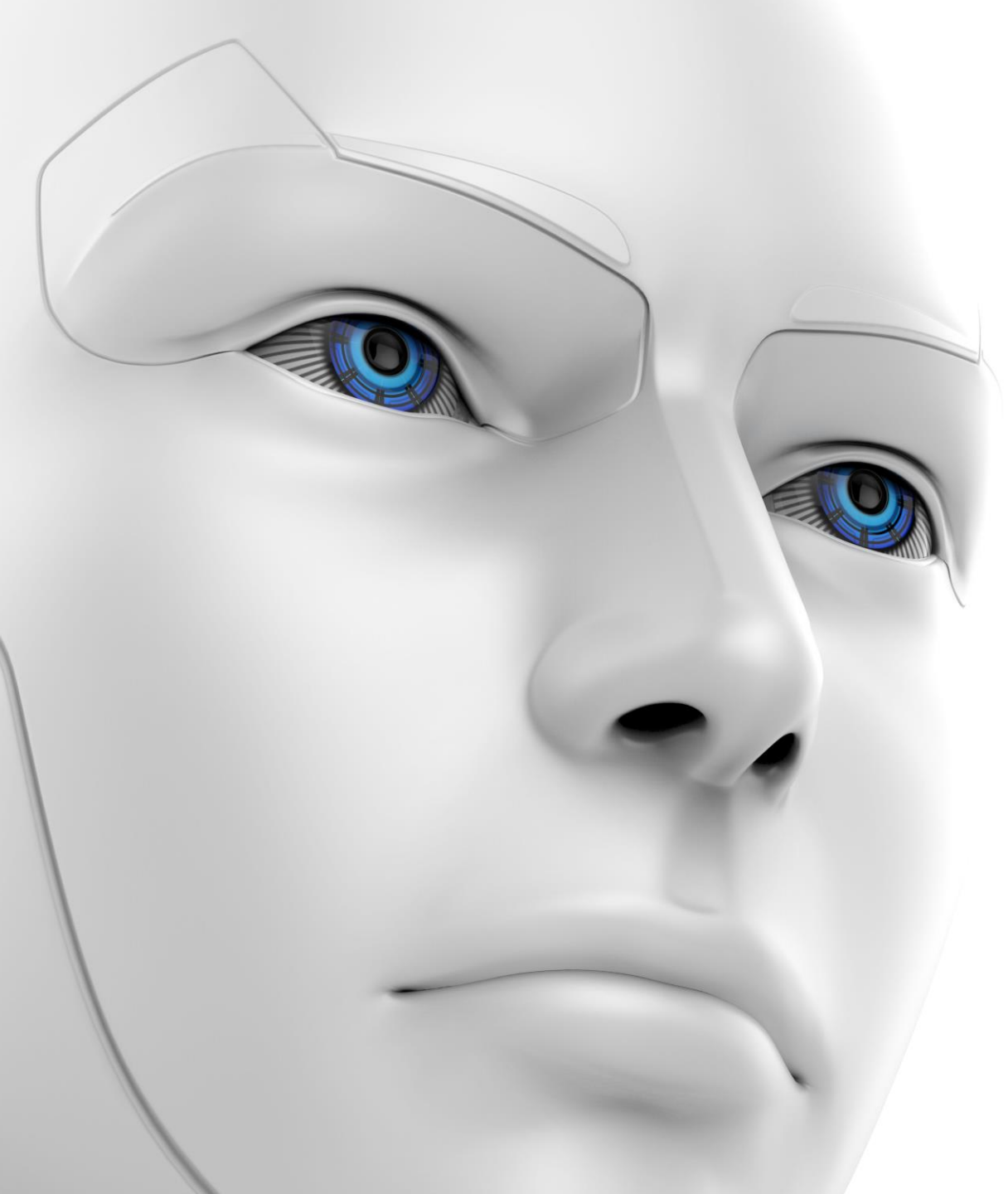


Dla osób głębiej wchodzących w temat AI, z czasem staje się coraz bardziej oczywiste, że nie istnieje jedna, wszechogarniająca definicja sztucznej inteligencji, ale raczej szeroki zakres definicji, które próbują uchwycić, co tak naprawdę ten termin oznacza. Im hasło AI staje się popularniejsze i marketingowo chwytliwe, tym więcej można odkryć sposobów jego użycia. Sensownych i bezsensownych. Nie dziwny się zatem, kiedy hasło AI bywa używane w przypadku, gdy żadnej sztucznej inteligencji (czyli w praktyce uczących się samodzielnie algorytmów) tam nie ma.

A futuristic robot head is shown in grayscale on the left side of the image. The robot has a metallic, reflective surface and a human-like face. A large, semi-transparent yellow rectangle is overlaid on the right side of the robot's face. In the center of this yellow rectangle is a white square containing the number '3.' in a bold, black, sans-serif font. The background is a blurred, grayscale image of a complex, geometric structure, possibly a futuristic building or a network of pipes.

3.

**KRÓTKA
HISTORIA AI**



Samo pojęcie AI wydaje się brzmieć dosyć nowocześnie.

Natomiast jest to idea rozwijająca się **już 80 lat**. Za początek AI uważa się **rok 1942** roku, kiedy **Isaac Asimov**, autor książek science fiction, ogłosił trzy prawa robotyki. W tamtych czasach AI kojarzyło się przede wszystkim z humanoidalnymi robotami, które będą towarzyszyć nam w życiu i prowadzić z nami dyskusje filozoficzne. Żeby ochronić nas przed możliwymi konsekwencjami zaistnienia nowej świadomości, **Asimov od razu ustalił jasne zasady:**

1. Robot **nie może skrzywdzić** istoty ludzkiej lub – poprzez wstrzymanie się od działania – pozwolić, by stała jej się krzywda.
2. Robot **musi wykonywać rozkazy** wydawane mu przez istoty ludzkie, z wyjątkiem sytuacji, kiedy byłyby one sprzeczne z Prawem Pierwszym.
3. Robot **musi chronić swoje istnienie** dopóty, dopóki taka ochrona nie jest sprzeczna z Prawem Pierwszym lub Drugim.

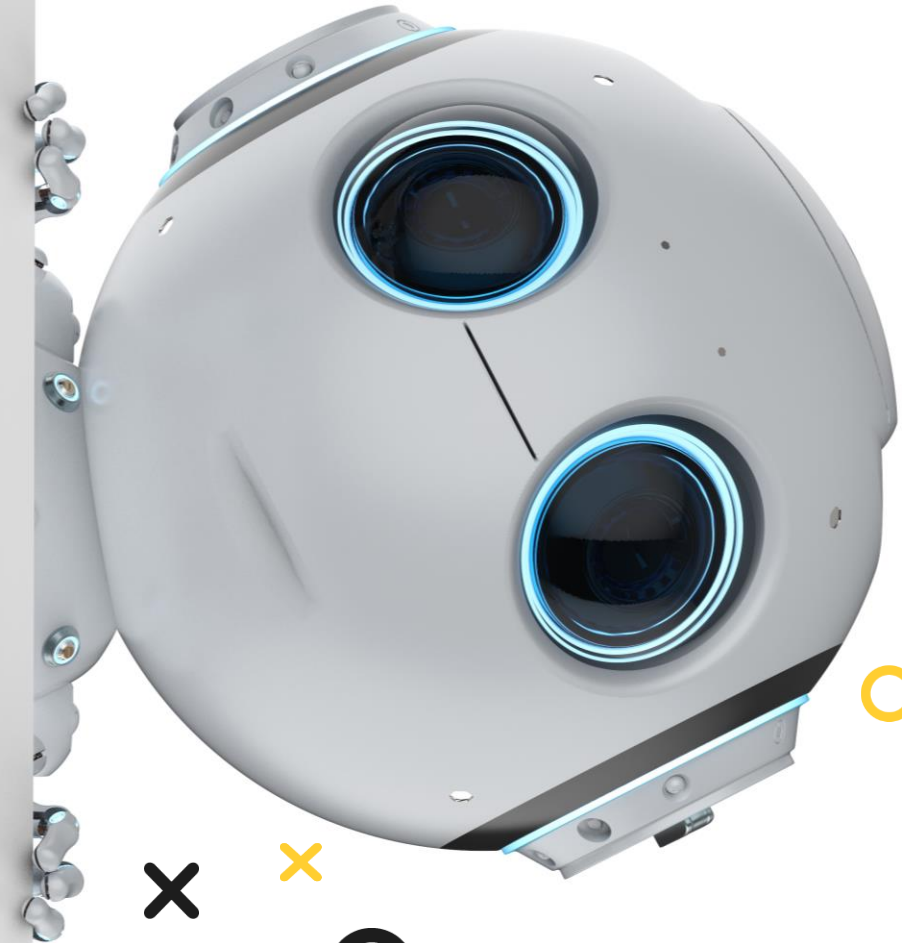
Z kolei w **1950** roku opublikowano **Test Alana Turinga**. Ten matematyk był zainteresowany zagadnieniem, jak sprawdzić, czy AI działa prawidłowo. Uznał, że jeśli rozmówca kierujący się tylko głosem nie rozpozna, czy ma do czynienia z człowiekiem czy maszyną, możemy mówić o prawdziwej sztucznej inteligencji. Zrównał zatem zdolności komunikacyjne, związane z mową i porozumiewaniem się za istotę AI. Dodajmy, że w tych czasach to były dyskusje stricte filozoficzne i tworzenie science fiction, żadnego AI w praktyce jeszcze nie było. **Koncepty rodziły się w głowach uczonych i pisarzy**, w większości przypadków mężczyzn w mocno średnim wieku.

Lata 1956–1973 uznaje się za złotą epokę AI. Świat naukowy zaczyna otrzymywać dotacje na realizację pierwszych projektów mających urzeczywistnić idee z poprzedniego okresu. W Dartmouth odbywają się pierwsze letnie naukowe warsztaty poświęcone AI. Miało to być wydarzeniem założycielskim sztucznej inteligencji jako dziedziny naukowej. Projekt trwał około sześciu do ośmiu tygodni i był zasadniczo rozszerzoną sesją burzy mózgów. **W 1968 powstał SHRDLU**, czyli system komputerowy służący do przetwarzania języka naturalnego, napisany w ramach doktoratu przez **Terry'ego Winograda** na Massachusetts Institute of Technology.

Co potrafił SHRDLU? Był to pierwszy AI, który wykonywał konkretne zadanie. Dzięki zainstalowanemu ramieniu układał klocki oraz odpowiadał na zadawane pytania na specjalnym monitorze. Po wydaniu komendy „podnieś czerwony klocek”, SHRDLU musiał znaleźć do niego drogę, podnieść go i dopiero wtedy potwierdzał na ekranie „ok”.

1969 to rok premiery robota Shakey, a w latach 70-tych powstały **MYCIN i Prolog**. Celem systemu Mycin było zdiagnozowanie bakteryjnej choroby krwi i zaproponowanie terapii. MYCIN miał wprowadzony zestaw reguł i bazę danych utworzoną przez konsylium lekarskie. Prolog z kolei to jeden z najpopularniejszych języków programowania logicznego, do automatycznej analizy języków naturalnych jako język ogólnego zastosowania.

Po okresie optymizmu i wiary w sukces AI nastąpiło załamanie. **Lata 1973–1980 określa się jako AI Winter**. Inwestorzy zwątpili we włożone w sztuczną inteligencję pieniądze i zwrot z inwestycji. „Artificial Intelligence: A General Survey”, powszechnie znany jako **raport Lighthilla**, został opublikowany w czasopiśmie Artificial Intelligence w 1973 roku. Oceniał stan badań w dziedzinie AI. Raport podał bardzo pesymistyczne prognozy, stwierdzając, że „w żadnej części tej dziedziny odkrycia dokonane do tej pory nie wywarły większego wpływu, jaki obiecywano”. Kolejny znaczący cios przyszedł w 1980 rok. **Zadał go filozof John Searl**. W historii zostanie zapamiętany jako **Argument Chińskiego Pokoju** (Chinese Room Argument). Searl udowadniał, że nawet jeśli komputer wydaje się, że posiada rozum, nie znaczy to, że ma rozum. Wykonywanie określonych zadań nie musi opierać się na rozumieniu ich przez wykonawcę. Przez 7 kolejnych lat w branży AI wiało chłodem.



Na szczęście w 1980 nastąpiła „Nowa Era AI” i nadzieja wróciła do branży. To czas projektów między innymi **CYC** (1984), **Nouvelle AI** (1985) czy **Deep Blue** (1997). Wyjaśniając po kolei: CYC (od ang. 'encyclopedia') został założony i prowadzony przez Douglasa Lenata. Składał się z bazy wiedzy zawierającej ręcznie zakodowane „zdroworozsądkowe” fakty oraz silnik wnioskowania, który pozwalał wydedukować dalsze fakty. Funkcjonuje do tej pory jako AI w wersji komercyjnej i oczywiście po kolejnych upgrade'ach. Natomiast Nouvelle (AI) to podejście do sztucznej inteligencji zapoczątkowane przez **Rodneya Brooksa**, który wówczas pracował w laboratorium sztucznej inteligencji MIT.

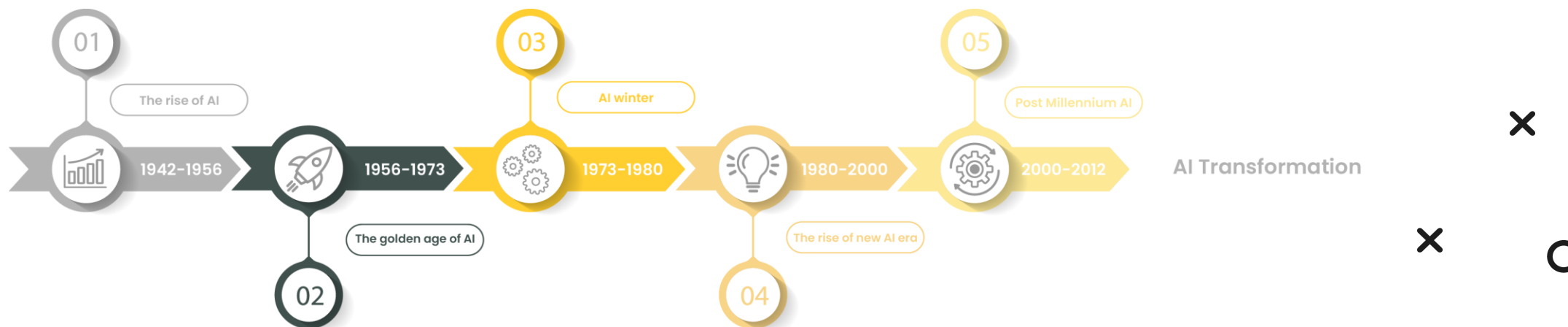
Co różniło Nouvelle AI od innych podejść? Celem tego projektu było stworzenie robotów o poziomie inteligencji podobnym do owadów. Uważano, że inteligencja może wyłonić się organicznie z prostych zachowań, gdy wchodzi w interakcję z „**rzeczywistym światem**”. To dosyć rewolucyjne podejście, ponieważ dotąd AI uczyły się na sztucznie 'skonstruowanych, symbolicznych światach' a nie w świecie rzeczywistym. Niezwykle ciekawym projektem był **Deep Blue**. Tak, to AI wygrało mecz szachów z mistrzem Kasparowem. To także spełnione marzenie firmy IBM o stworzeniu superkomputera. Zwycięstwo Deep Blue zostało uznane za kamień milowy w historii sztucznej inteligencji i było tematem kilku książek i filmów.



Kolejnym etapem rozwoju AI był okres nazywany jako **Post-Millennium AI**, trwający w latach 2000–2012. Sztuczna Inteligencja zaczęła znajdować praktyczne i bardziej opłacalne biznesowo rozwiązania. W 2002 roku powstał **odkurzacz Roomba**, projekty wojskowe, jak pomagający nosić żołnierzom ciężary robot **Big Dog**, natomiast w 2010 roku pojawiła się **Siri** jako asystent głosowy. **Rok później świat poznał Watsona**, czyli sztuczną inteligencję z firmy IBM. Potentat obiecywał, że ich produkt przekształci branżę i będzie generować przychód dla firm. Watson został początkowo opracowany tak, aby odpowiadać na pytania w znanym amerykańskim teleturnieju „Jeopardy!”. Zdobyl nawet nagrodę za pierwsze miejsce w wysokości 1 miliona dolarów, po zwycięstwie nad ludzkimi rywalami. Następnie IBM ogłosił, że system będzie dalej rozwijany w kierunku branży medycznej jako pomoc w **stawianiu diagnoz**. System był używany jako wsparcie np. w leczeniu raka płuc w Memorial Sloan Kettering Cancer Center w Nowym Jorku.

Lata 2000–2012 to też okres, w którym ruszyło **laboratorium badawcze DeepMind** pracujące głównie nad sieciami neuronowymi ANN (Artificial Neural Network). DeepMind uczyło się grać w gry wideo i trafiło na pierwsze strony gazet w 2016 roku po tym, jak jego program AlphaGo pokonał profesjonalnego gracza **Lee Sedola**, mistrza świata w grze w „Go”. Ogólniejszy program, AlphaZero, pokonał najpotężniejsze programy grające w szachy i shogi (japońskie szachy) po kilku dniach gry przeciwko sobie. Sukcesem skomplikowanych prac było finalnie stworzenie komputera naśladowującego pamięć krótkotrwałą ludzkiego mózgu. DeepMind został przejęty przez Google w 2014 roku i rozwija się dalej. W 2020 roku DeepMind poczynił znaczne postępy w problemie fałdowania białek dzięki AlphaFold, a w bazie danych zostało opublikowanych ponad 200 milionów przewidywanych struktur, reprezentujących praktycznie wszystkie znane białka. W 2022 roku DeepMind ogłosił, że rozwija język wizualny (VLM) o nazwie **Flamingo** oraz opracowuje **DeepNash**, czyli wieloagentowy system uczenia się bez modelowania.

W jakiej epoce AI dzisiaj żyjemy po 2012 roku? Nauka nie podaje jeszcze nazwy. Postulujemy nazwę „AI transformation” albo „quiet explosion”. Dodajmy, praktycznie w każdej branży i dziedzinie życia. Prawo Amary mówi, że ludzie mają tendencję do przeceniania wpływu technologii na krótką metę, ale nie doceniają jej na dłuższą metę. My już na tej dłuższej mecie jako społeczeństwo jesteśmy.





4.

PRZEGLĄD TECHNIKI
ALGORYTMÓW

Filarem sztucznej inteligencji, jak już wiemy, jest **uczenie maszynowe** (choć nie jedynym, ale bardzo znaczącym). Uczenie maszynowe (ang. machine learning, ML) to grupa algorytmów, które pozwalają komputerom uczyć się na podstawie danych i poprawiać swoją dokładność w czasie. Jak już wiemy z rozdziału „Definicje AI” machine learning możemy podzielić na następujące rodzaje:

- **nadzorowane** (ang. supervised learning),
- **nienadzorowane** (ang. unsupervised learning),
- **ze wzmocnieniem** (ang. reinforcement learning),
- **uczenie głębokie** (ang. deep learning).

Poszczególne typy zostaną szczegółowo opisane poniżej, jednak na uwagę zasługuje przede wszystkim **deep learning**. Pomimo tego, że jest podtypem uczenia maszynowego, ze względu na swoją szczególną efektywność „opuścił” rodzinne gniazdo uczenia maszynowego i „wyrabia” sobie własną drogę życiową i nazwisko. Warto dodać, że przy ogromnym zapotrzebowaniu na niezbędną ilość danych do uczenia. Ta ekspansja i samodzielność deep learningu spowodowała pewną nieścisłość definicyjną, tzn. **uczenie maszynowe** utożsamiane jest z uczeniem nadzorowanym i nienadzorowanym, natomiast **uczenie głębokie** traktowane jest osobno (mimo, że jest podtypem/dzieckiem uczenia maszynowego). Nie przeszkadza to rozwojowi zarówno klasycznym metodom uczenia maszynowego, jak i głębokiemu uczeniu.



4.1

NADZOROWANE UCZENIE MASZYNOWE

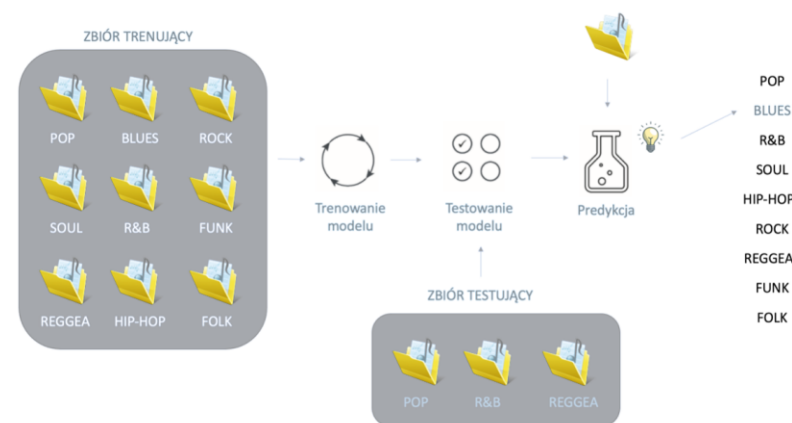
Uczenie takie polega na wskazaniu algorytmowi ciągu danych, np. cech danego obiektu wraz z klasą obiektu lub wartością liczbową (np. ceną). **Im więcej takich danych, tym lepiej.** Po procesie uczenia wystarczy podać zestaw cech bez klasy obiektu lub wartości liczbowej, a algorytm sam wybierze kategorię (klasę) obiektu lub jego wartość. Przewidywanie kategorii klasy nazywamy **klasyfikacją**, a szacowanie wartości – **regresją**.

Spróbujmy to jeszcze lepiej zobrazować na bardzo sugestywnym przykładzie. Zaczniemy od klasyfikacji. Wyobraźmy sobie, że chcemy nauczyć dziecko (a może już nastolatka) rozróżniać style muzyczne. Dziecko/nastolatek jest więc naszym modelem AI, który chcemy odpowiednio wytrenować. Dane wejściowe to zbiór utworów muzycznych przypisanych do jednej kategorii (klasy), np. ROCK, POP, BLUES itd. Czyli obiekty to utwory muzyczne, klasy/kategorie to style muzyczne. Jak wiadomo, każdy obiekt (tj. utwór muzyczny) posiada atrybuty (cechy), np. długość w sekundach, skład zespołu wraz z rodzajami instrumentów, tempo, tekst itp. W zbiorze danych (dataset) mamy ponad 25 000 piosenek ze wszystkimi atrybutami i stylem muzycznym (czyli klasą). Ponieważ klasa jest pewną etykietą danego obiektu, taki zbiór danych jest często określany jako etykietowany (labeled dataset). Niestety, uczenie nadzorowane wymaga etykietowanych zbiorów danych, czyli obiektów opisanych przez atrybuty z przypisanymi klasami. Teraz dzielimy nasz zbiór danych na dane treningowe i testowe, np. 80% danych (20 000 ścieżek) posłuży nam do trenowania modelu, a 20% (5 000 ścieżek) do testowania modelu. Drukujemy więc nasz zestaw danych i wręczamy go dziecku do nauczania. Czyta ono ścieżkę po ścieżce i analizuje, robiąc sobie notatki. Szuka wzorców w danych, np. piosenki PUNK są bardzo krótkie, często występują w nich małe składy i rzadko natomiast instrumenty elektroniczne, tempo jest szybkie, a w tekstach często występuje fraza NO FUTURE. Z kolei w DISCO POLO brakuje gitar i perkusji, a syntezatory są często obecne. Słowa piosenek są zawsze po polsku i jest w nich dużo frywolnych zwrotów wraz z częstymi rymami. Po dwóch miesiącach nasze dziecko ukończyło trening, ma sto stron notatek i wyszukało wiele wzorców. Nadszedł czas na przetestowanie modelu.

Dajemy więc zestaw testowy (5 tys. piosenek), ukrywamy styl muzyczny, a nasze dziecko musi przewidzieć odpowiedni styl muzyczny na podstawie notatek i wzorców. Utwór po utworze z zestawów testowych. Gdy skończy, sprawdzimy i obliczymy dokładność modelu według wzoru:

$$\text{dokładność} = \frac{\text{liczba poprawnych predykcji}}{\text{liczba utworów w zbiorze testowym}}$$

Co jeśli dokładność jest niezadowalająca? Po pierwsze, może za mało danych? Po drugie może dane są słabej jakości? Po trzecie może hmmm... coś jest nie tak z dzieckiem. Ponieważ nasz model przypisywał obiekt do stylu czyli danej kategoriycznej (kategorii) to możemy nazwać go klasyfikatorem (zgadywanie kategorii/klasy w naszym przypadku stylu muzycznego).



Uczenie nadzorowane na przykładzie rozpoznawania stylu muzyki

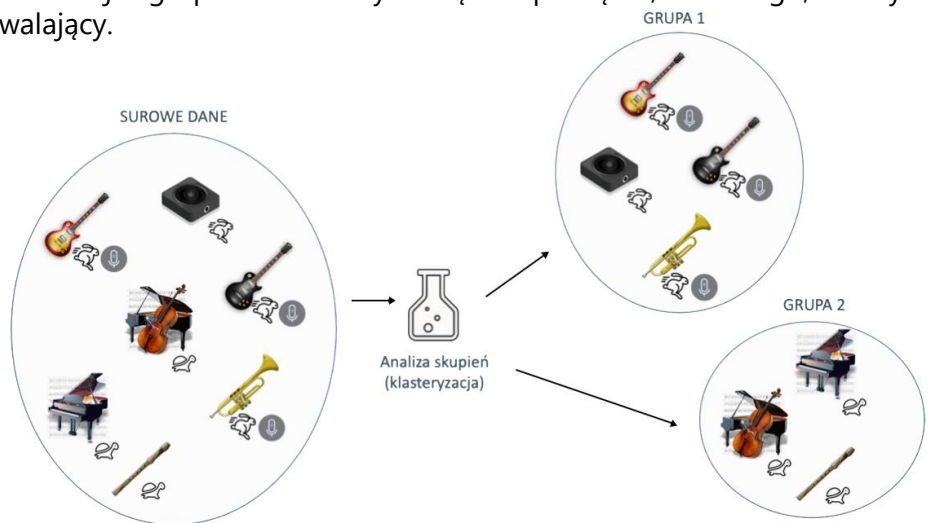
Do czego możemy wykorzystać takie uczenie nadzorowane? Do identyfikacji spamu, analizy emocji, rozpoznania języka, określania ryzyka kredytowego, kategoryzacji dokumentów, określenia wartości mieszkania, samochodu (automatyczna wycena), oszacowania czy klient kupi dany produkt czy nie itp.

4.2

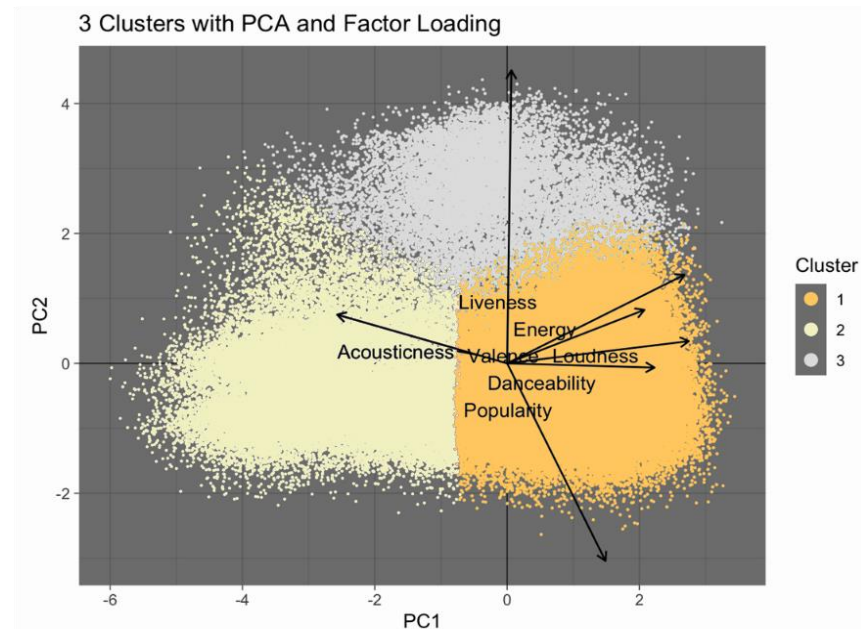
NIENADZOROWANE UCZENIE MASZYNOWE

W przypadku uczenia bez nadzoru, w tzw. **danych treningowych**, mamy tylko cechy obiektów. Możemy więc jedynie „uporządkować” dane, podzielić je na grupy podobnych obiektów (klasteryzacja, analiza skupień – **clustering**) lub zmniejszyć ich wymiary (redukcja wymiarów, dimension reduction), czyli dokonać selekcji cech pod kątem dalszego uczenia maszynowego, łatwiejszej wizualizacji lub zmniejszenia wykorzystania zasobów komputera (np. pamięci RAM).

Jak w praktyce wygląda na przykład analiza skupień (clustering)? Znowu wróćmy do naszego przykładu muzycznego i naszego nastoletniego geniusza. Otóż teraz wyobraźmy sobie, że utwory muzyczne posiadają opisy/cechy tj. np. długość w sekundach, skład zespołu wraz z rodzajami instrumentów, tempo, tekst, ale nic nie wiemy o stylu muzycznym. Dlatego niech nasz dzieciak przejrzy wszystko i podzieli na X grup (np. dziesięć) składających się z podobnych utworów. Nie pokazaliśmy mu przykładowych grup, nie powiedzieliśmy na czym polega dokładnie podobieństwo, po prostu uczenie bez nadzoru – kombinuj. No i mały geniusz biedny siedzi i próbuje, a może oddzielną grupę zrobię utwory szybkie z wokalistką śpiewającą po polsku. Hmm... słabo tylko 20 utworów, a w innych grupach po tysiące. Zatem może inaczej i grupowanie zaczyna się od początku, tak długo, aż wynik będzie zadowalający.



Uczenie nienadzorowane na przykładzie klasteryzacji utworów muzycznych



Klastrowanie utworów muzycznych z Spotify (algorytm K-Means i PCA, uczenie nienadzorowane)

Źródło: <https://www.rpubs.com/ajibudiarto/spotify> [dostęp: 25.08.2022]

Zapewne u wielu niedowiarków zaczyna pojawiać się pytanie, a po co nam grupy (skupiska) utworów muzycznych, zwierząt, ludzi, klientów, skoro są nienazwane? Jednak czy wszystko musimy nazywać? Wyobraźmy sobie, że do algorytmu przekazujemy np. zachowania zakupowe określonych użytkowników, np. informacje o tym, jakie produkty kupują, a jakie tylko przeglądają. Algorytm automatycznie dzieli użytkowników na podaną liczbę grup na podstawie ich podobieństwa. Pamiętajmy, że np. podobni klienci zachowują się podobnie, mają podobne potrzeby zatem czy potrzebujemy ich nazwać czy wystarczy prostą statystyką wyciągnąć z grupy najczęściej kupowane produkty i zarekomendować reszcie grupy. **Dzięki klastrowaniu/grupowaniu można bardzo szybko stworzyć spersonalizowane biuletyny, rankingi i wygląd strony głównej bez potrzeby nazywania danej grupy.**

4.3

UCZENIE MASZYNOWE ZE WZMOCNIENIEM

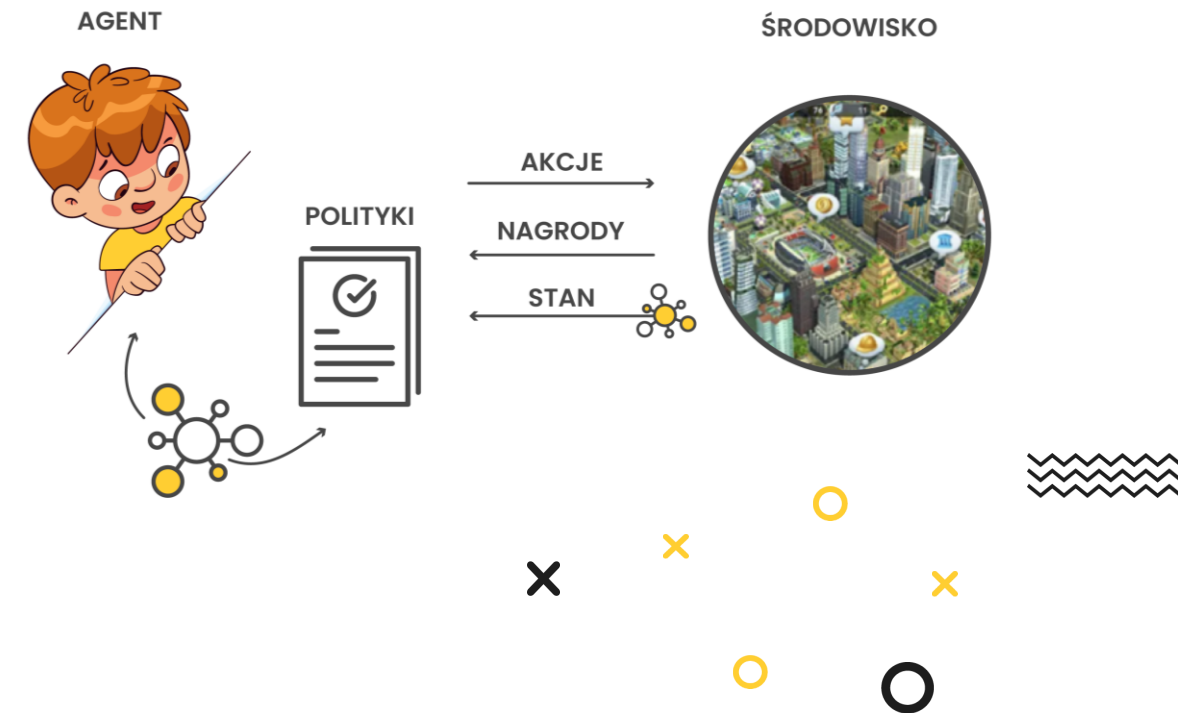
Reinforcement Learning (RL) znalazło zastosowanie głównie w przemyśle (robotyka), motoryzacji (samochody autonomiczne), domach maklerskich oraz w tzw. **GameDev** (branża gier komputerowych), czyli wszędzie tam, gdzie uczenie decyzyjne odbywa się metodą prób i błędów. No i przede wszystkim, między innymi, również w ChatGTP.

Jest to jeden z najmniej popularnych typów uczenia maszynowego, a jego działanie najbardziej różni się od standardowych algorytmów z rodziny maszynowych procesów uczenia. Po pierwsze, brakuje w nim danych treningowych. Pojawiają się tu jednak nowe terminy tj. środowisko (environment), z którego algorytm może automatycznie zbierać dane. Dane te są przekazywane do bufora i służą do trenowania agentów. Głównym celem agenta jest nauczenie się jak najlepiej współdziałać ze środowiskiem.

Może uda nam się też wyjaśnić tę zasadę na przykładzie dziecka i gry komputerowej. Wyobraźmy sobie więc, że kupiliśmy prostą grę strategiczną, w której budujemy i rozwijamy miasta. Oczywiście mogliśmy dokładnie przeczytać instrukcję, ale może jednak metodą prób i błędów nauczymy się zasad gry. Nasze dziecko będzie więc, zgodnie z nomenklaturą RL – agentem, podczas gdy gra – środowiskiem. W początkowej fazie zbierania danych nasz agent rozegra kilka tur w grze, bardzo często opartych na losowych działaniach. Tu coś zbuduje lub zburzy, tam podniesie podatki. Jeśli w danej akcji otrzyma nagrodę (przybędzie punktów lub trochę monet), to skrupulatnie odnotowuje to w notatkach (tzw. polityka) wraz ze stanem środowiska (gry). Oczywiście tych tur (rozgrywek) nasz agent musi trochę pograć, dodatkowo czasem musi dobrze przeanalizować ściągę (politykę), czy np. nie warto pominąć kilku małych nagród, aby zebrać jedną i porządną (akcje długoterminowe). W praktyce więc do tych procesów uczenia się (analizy obserwacji) czy sposobów kolekcjonowania danych, stosuje się inne algorytmy ML (np. sieci neuronowe).



Uczenie maszynowe ze wzmocnieniem na przykładzie odkrywania zasad gry strategicznej



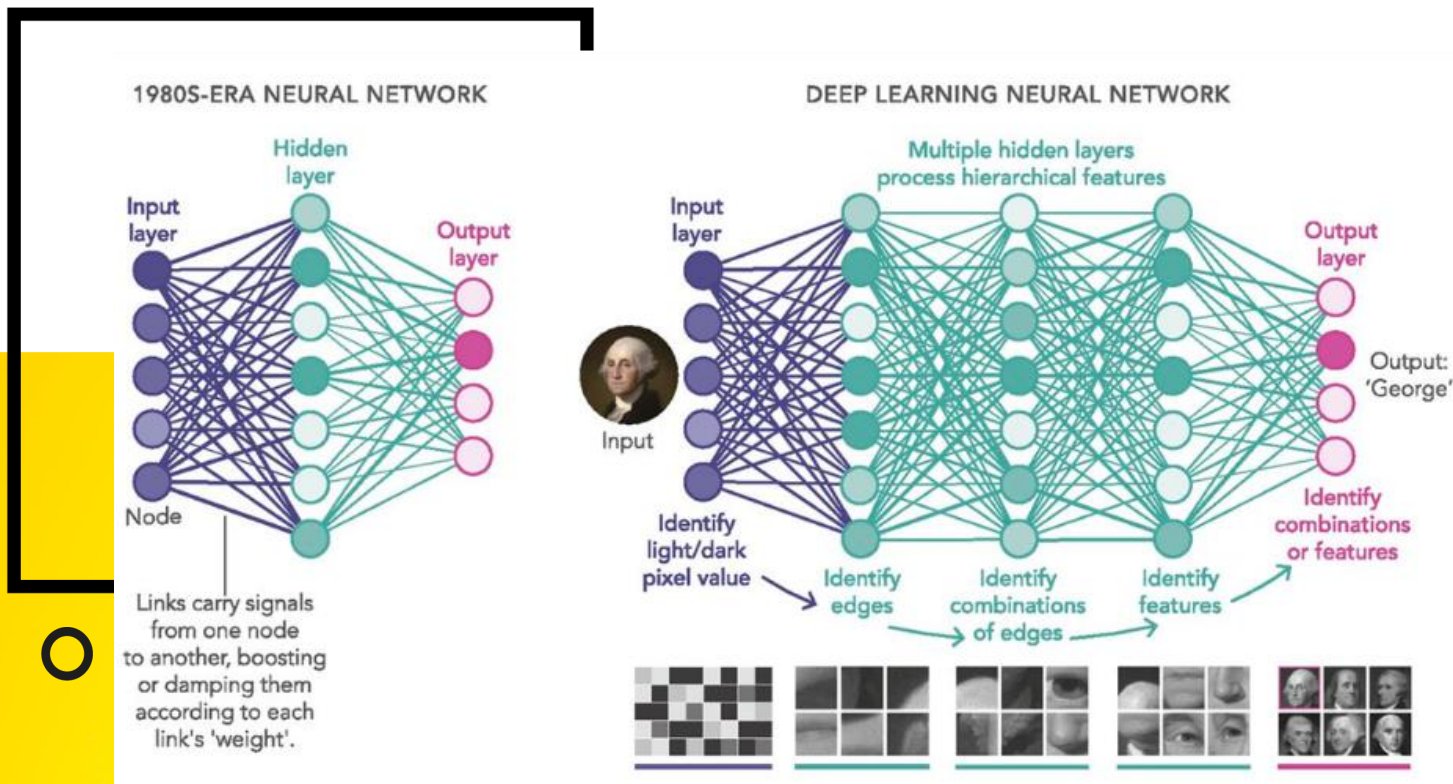


4.4

**GŁĘBOKIE
UCZENIE**

Deep learning (głębokie uczenie) to podtyp uczenia maszynowego, który jest intensywnie rozwijany. Trzeba przyznać, z bardzo dobrymi wynikami. W tym rodzaju uczenia można dostrzec istotne podobieństwa do tradycyjnych metod uczenia maszynowego, np. w obszarze wysokopoziomowych zasad działania (takich jak proces szkolenia/uczenia). Istnieją jednak również istotne różnice, takie jak wymóg dużych zbiorów danych i długi czas uczenia bez konieczności ręcznej ekstrakcji cech.

Nazwa grupy algorytmów, czyli głębokie uczenie (deep learning), pochodzi od wykorzystania struktur **sztucznych sieci neuronowych**, które mogą składać się z wielu warstw (wejściowej, ukrytej i wyjściowej), przez co proces uczenia sprawia charakter bardzo głębokiego. Należy wspomnieć, że nie każde użycie sieci neuronowej jest już głębokim uczeniem, a wszystko zależy od liczby zastosowanych warstw. Tak więc, pomimo funkcjonalnego podobieństwa do nadzorowanego lub nienadzorowanego (klasycznego) uczenia maszynowego, podstawowa różnica polega na tym, że deep learning musi wykorzystywać sieci neuronowe z odpowiednią liczbą warstw (min. 3).



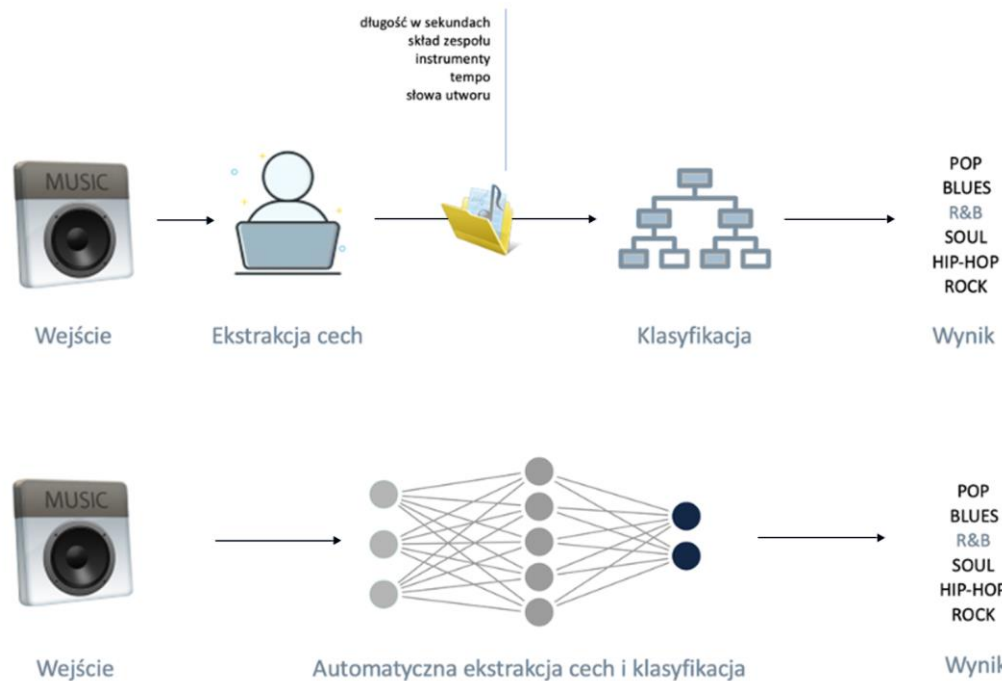
Prosta sieć neuronowa vs algorytm głębokiego uczenia. Liczba warstw ma znaczenie

Źródło: <https://www.pnas.org/doi/10.1073/pnas.1821594116>
[dostęp: 25.08.2022]

Co zatem zapewnia zastosowanie „potężniejszych” algorytmów wzorowanych na ludzkim układzie nerwowym?

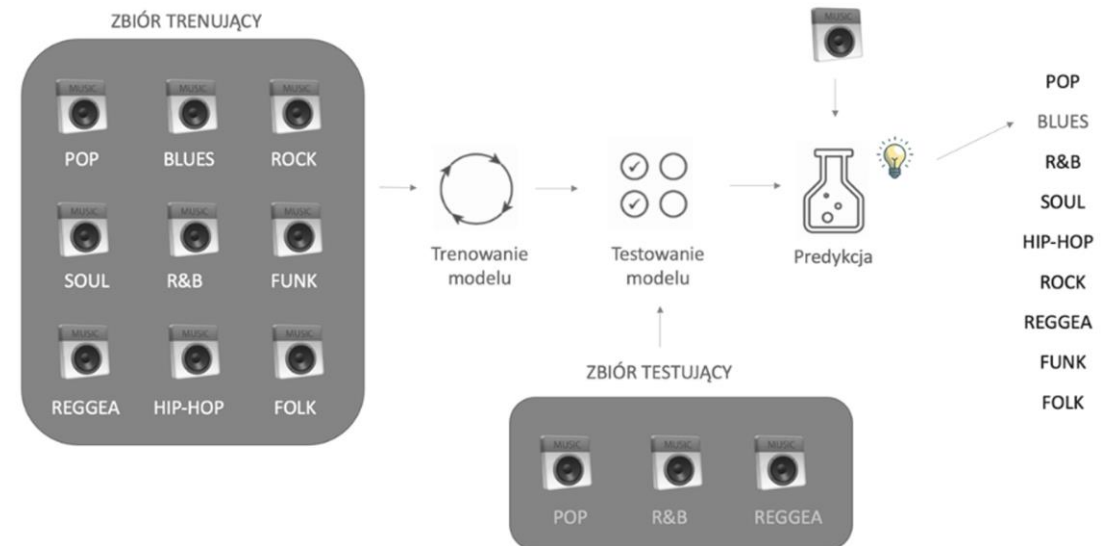
Na przykład wspomniany brak ręcznej ekstrakcji cech czy lepszą dokładność predykcji. Dziecięcy przykład? Czemu nie.

Wróćmy jeszcze raz do naszego muzycznego przypadku i nadzorowanego (klasycznego) uczenia. Na wejściu nasz geniusz miał piosenki „opisane” za pomocą cech np. instrumentarium, tekst, długość piosenki, tempo, czyli ktoś (i nie był to algorytm ML tylko fizyczna osoba) musiał nie tylko przypisać style do piosenek, ale również określić te cechy. Natomiast w deep learningu mamy o wiele **więcej możliwości**. Więc czy nie wystarczy po prostu zagrać daną piosenkę i pozwolić naszemu dziecku posłuchać, aby dowiedziało się jak piosenka brzmi w aranżacji na dwie gitary, bas, perkusję? Z drugiej strony, jak brzmią piosenki dodatkowo na przykład z sekcją dętą? Po **19 234** przesłuchaniu na pewno nauczy się rozpoznawać pewne wzorce w brzmieniu utworu.



W przypadku głębokiego uczenia najczęściej algorytmy samodzielnie radzą sobie z ekstrakcją cech

Zatem automatyczna ekstrakcja cech w przypadku deep learningu jest już obecna, więc najczęściej jako dane wejściowe dla algorytmów mamy obrazy, dźwięk, wideo, długie teksty itp. Natomiast sam proces **wysokopoziomowo** przypomina klasyczne uczenie maszynowe. Przykładowo, proces klasyfikacji naszych utworów wygląda tak samo, tylko na wejściu nie są to np. pliki tekstowe z opisem piosenki, a sama piosenka.



Głębokie uczenie (przykładowy proces klasyfikacji) na przykładzie rozpoznawania stylu muzyki

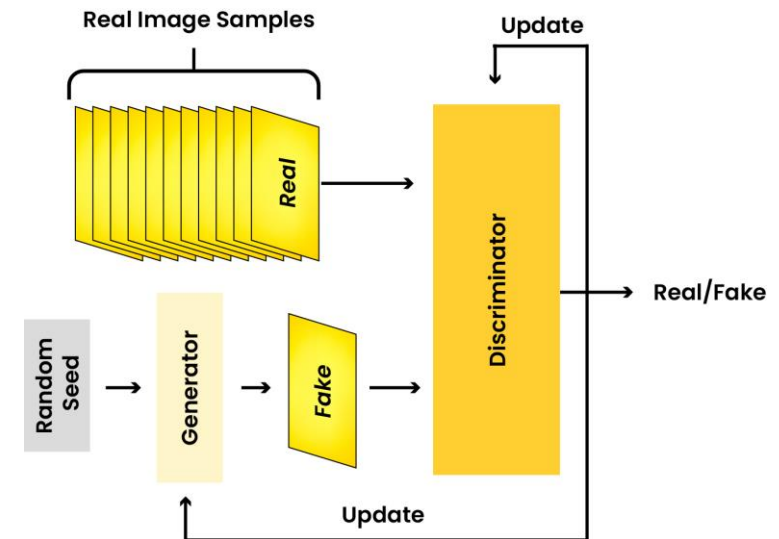
Zanim jednak wyniknie wnioski, że klasyczne algorytmy uczenia nadzorowanego niewiele różnią się od deep learningu, przyjrzyjmy się pojęciu tzw. architektury. Wiemy, że aby model należał do rodziny głębokiego uczenia, **musi zawierać sieć neuronową** z co najmniej trzema warstwami. Co nam jednak szkodzi użyć nie tylko zestawu większej ilości warstw, ale także kilku różnych sieci? Albo np. obok sieci neuronowej użyć innego bytu jak algorytmu, naśladującego proces poznawczy organizmu ludzkiego?

GAN – Generative Adversarial Network

Rozwój Internetu oraz wyposażenie każdego telefonu w wysokiej jakości aparat fotograficzny lub kamerę spowodował, że społeczeństwo masowo kręci filmy, robi zdjęcia sobie, rodzinie, znajomym. Nawet jedzenie i paragony stały się ulubionym tematem zdjęć. **Już w 2013 roku użytkownicy zamieszczali w sieci 500 milionów zdjęć dziennie.** Aż się prosiło, by zacząć przetwarzać te obrazy algorytmami uczenia głębokiego. Segregacja (klasyfikacja), wykrywanie emocji (np. uśmiechów) czy znajomych twarzy, wyszukiwanie wizualne (np. powiedz mi algorytmem, co ma na sobie celebryta) to problemy, z którymi szybko poradziło sobie głębokie uczenie za pomocą np. konwolucyjnych sieci neuronowych. Użytkownicy chcieli jednak czegoś więcej, np. jak będę wyglądał za 30 lat albo umieszczenia zdjęcia danej osoby we fragmencie przebojowego filmu. Z drugiej strony naukowcy mieli dostęp do ogromnych zasobów zdjęć, na bazie których można było nauczyć modele, które nie tylko świetnie radzą sobie z klasyfikacją czy identyfikacją obiektów, ale są też kreatywne i wygenerują ze zdjęcia co tylko chcemy.

W 2014 roku, podczas świętowania doktoratu w montrealским barze, grupa przyjaciół poprosiła naukowca **Iana Goodfellow'a** o dyskusję problemu braku wyobraźni w algorytmach AI. **Goodfellow**, przy kuflu piwa, zaproponował prostą koncepcję połączenia dwóch sieci neuronowych, które konkurują ze sobą (w formie gry), tzn. jedna z nich – kreator generuje np. fotorealistyczne obrazy, a druga – dyskryminator próbuje rozpoznać, czy obraz wygląda już prawdziwie, czy dalej „zalatuje” sztucznością. Oba modele poprawiają swoje umiejętności podczas procesu uczenia się. Generator generuje coraz lepsze obrazki, a dyskryminator coraz lepiej rozróżnia obrazy fałszywe od prawdziwych. Podobno większość uczestników była sceptyczna, ale Ian nie poddał się i z dnia na dzień wdrożył rozwiązanie, nazywając je **generatywnymi sieciami współzawodniczącymi (GAN)**.

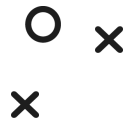
Dzięki GAN-om **Ian Goodfellow** stał się legendą AI i trafił do pracy w Google (obecnie pracuje w Apple). Z kolei sieci generatywne z jednej strony stały się „silnikiem” imponujących aplikacji generujących np. twarze nieistniejących osób, wygląd ludzi w przyszłości czy obrazy w danym stylu malarskim itp. To nie koniec zastosowań sieci współzawodniczących, gdyż znalazły one swoje miejsce np. w medycynie. Dzięki nim można np. realistycznie „rozmnożyć” zdjęcie rentgenowskie i na nich trenować/testować inne algorytmy, np. do diagnozowania chorób.



Architektura GAN

Źródło: https://semiengineering.com/knowledge_centers/artificial-intelligence/neural-networks/generative-adversarial-network-gan/ [dostęp: 25.08.2022]

Istnieje jednak ciemna strona architektury GAN przy generowaniu takich fotorealistycznych, a jednocześnie nieistniejących obrazów tj. deep fake. Nie trzeba nikomu tłumaczyć, że np. kompromitujące, a jednocześnie fałszywe wideo może zmienić wynik wyborów, pozbawić kogoś pracy lub niesłusznie oskarżyć.



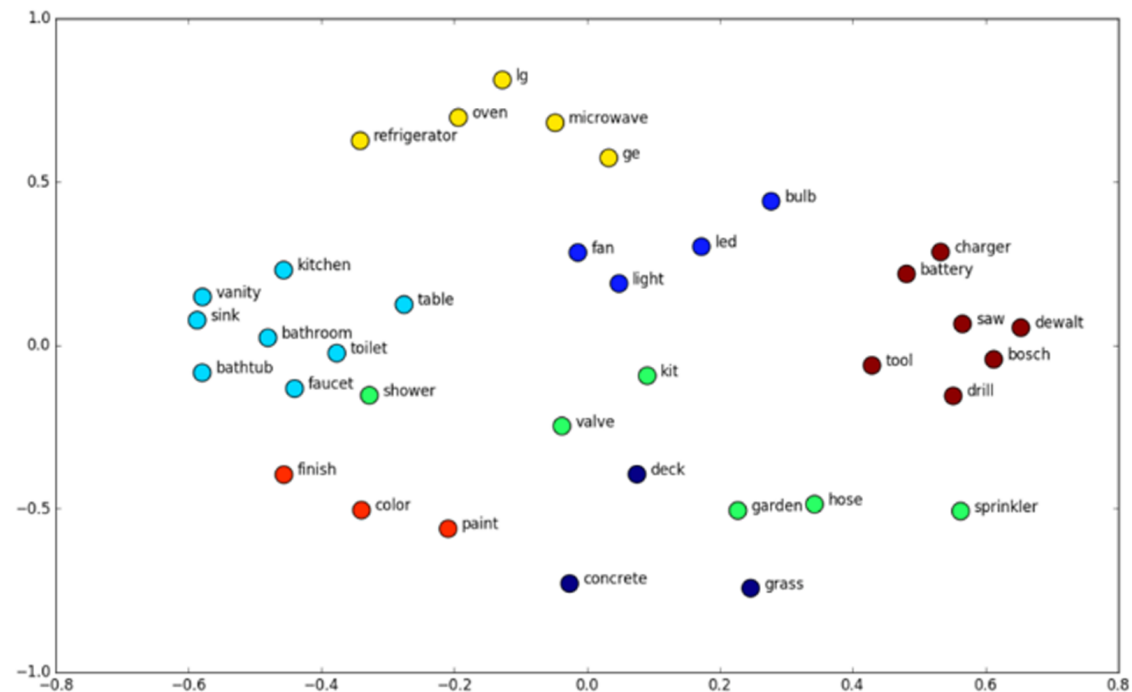
Transformery

Większość informacji i wiedzy przechowywana jest w tekstach (książki, strony internetowe). Dodatkowo coraz więcej tekstów powstaje każdego dnia podczas komunikacji interpersonalnej (maile, SMS-y, czaty). Potrzebujemy coraz wydajniejszych algorytmów/modeli do przetwarzania i generowania języka naturalnego. Jak wiemy tekst składa się ze zdań, a te zdania składają się ze słów o odpowiedniej kolejności, które często mają znaczenie (sekwencje). No i tu zaczyna się problem, bo wszelkie klasyczne algorytmy uczenia maszynowego przetwarzające właśnie język naturalny nie traktowały sekwencji jako całości, a bardziej zmieniały tekst na listę słów z odpowiednimi wagami. Czasem „zapominając” o kontekście, sekwencji i innych istotnych elementach.

Sieci neuronowe typu **RNN (rekurencyjne)** zaczęły obsługiwać dane sekwencyjne (teksty), analizując np. słowo po słowie. Jakość przetwarzania rosła, jednak do rewolucji było jeszcze trochę daleko. W 2017 roku zespół Google Brain stworzył koncepcję transformerów i opisał ją w dokumencie „**Attention Is All You Need**”.

Naukowcy głównie skupili się na mechanizmie uwagi, a dokładniej „samo-uwagi” (self-attention) wzorowanej na uwadze w procesach poznawczych i mocno wspierającym ją kontekście. Dzięki tej pracy naukowej powstała architektura **transformerów**, która „szturmem” wdarła się do laboratoriów BigTech i namieszała początkowo w **NLP**, a obecnie w niemal każdej dziedzinie przetwarzania danych.

Technicznie są to dalej modele korzystające z dobrodziejstwa głębokiego uczenia, ale ulepszone kilkoma dodatkami. Najważniejszym elementem jest „self-attention”, do tego traktowanie sekwencji jako całości oraz szereg technik już dobrze znanych z przetwarzania języka naturalnego, takie jak „word embedding” (osadzanie słów) czy też „positional encoding” (czyli znaczenie pozycji słowa w zdaniu).



Word embedding – wyuczona maszynowo reprezentacja słów (np. w postaci wektora), w której słowa o tym samym znaczeniu mają podobną reprezentację.

Źródło: <https://neptune.ai/blog/word-embeddings-guide> [dostęp: 25.08.2022]

Naukowcy zaczęli prezentować transformatory w postaci gotowych rozwiązań dostępnych w chmurze.

Jak grzyby po deszczu zaczęły powstać modele oparte na architekturze transformera tj.: BERT, T5, GPT itp. Najbardziej znanym przedstawicielem takich architektur jest GPT-3 (Generative Pre-trained Transformer 3) organizacji OpenAI. Organizacja OpenAI założona przez Elona Muska, a obecnie mocno wspierana przez Microsoft pracowała przez 12 miesięcy modelem GPT-3. Przy pracach zatrudnionych było ponad 30 naukowców z całego świata, niektóre z pierwszych wersji „trójki” miały ponad 175 miliardów parametrów. Do treningu użyto ogromnych ilości tekstu z różnych źródeł (książki, strony internetowe). Ilość była tak duża, że np. Wikipedia stanowiła tylko 3% całego zbioru.

GTP-3 służy głównie do generowania tekstu na podstawie kilku pierwszych słów. Potrafi to zrobić w taki sposób, że trudno odróżnić, czy tekst stworzył człowiek, czy maszyna. Dodatkowo możemy zażyczyć sobie, w jakim stylu ma go wygenerować, na przykład urzędowym czy szekspirowskim. Ale to nie koniec, bo transformator ten jest bardzo wszechstronny i może również: tłumaczyć tekst na inne języki, sprawdzać i poprawiać styl oraz gramatykę, wyciągać dane z tekstu, generować pytania do wywiadu, odpowiadać na pytania, wymyślać hasła reklamowe np. na podstawie parametrów produktu.



Przykładowe funkcjonalności (bo jest ich kilka razy więcej) GPT-3

Źródło: <https://beta.openai.com/> [dostęp: 25.08.2022]

The screenshot displays a grid of 18 application examples for GPT-3, each with a colored icon and a brief description:

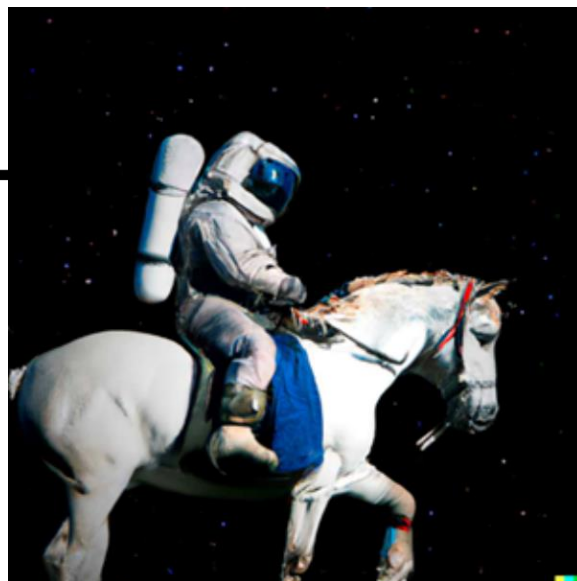
- Q&A**: Answer questions based on existing knowle...
- Summarize for a 2nd grader**: Translates difficult text into simpler concep...
- Text to command**: Translate text into programmatic commands.
- Natural language to Stripe API**: Create code to call the Stripe API using nat...
- Parse unstructured data**: Create tables from long form text
- Python to natural language**: Explain a piece of Python code in human un...
- Calculate Time Complexity**: Find the time complexity of a function.
- Advanced tweet classifier**: Advanced sentiment detection for a piece o...
- Keywords**: Extract keywords from a block of text.
- Ad from product description**: Turn a product description into ad copy.
- Grammar correction**: Corrects sentences into standard English.
- Natural language to OpenAI API**: Create code to call to the OpenAI API usin...
- English to other languages**: Translates English text into French, Spanish...
- SQL translate**: Translate natural language to SQL queries.
- Classification**: Classify items into categories via example.
- Movie to Emoji**: Convert movie titles into emoji.
- Translate programming languages**: Translate from one programming language ...
- Explain code**: Explain a complicated piece of code.
- Factual answering**: Guide the model towards factual answering ...
- Product name generator**: Create product names from examples word...

Transformatory mają jeszcze jedną ciekawą cechę, ponieważ można je „dostroić” do konkretnych zadań, o ile dane wejściowe i wyjściowe są zachowane w sekwencji i mamy wystarczająco dużo danych do treningu. Tak więc zespół na podstawie wyników prac nad GPT-3 stworzył model CODEX do automatycznego generowania kodu źródłowego programów lub DALL-E czyli generowanie obrazów z tekstu.



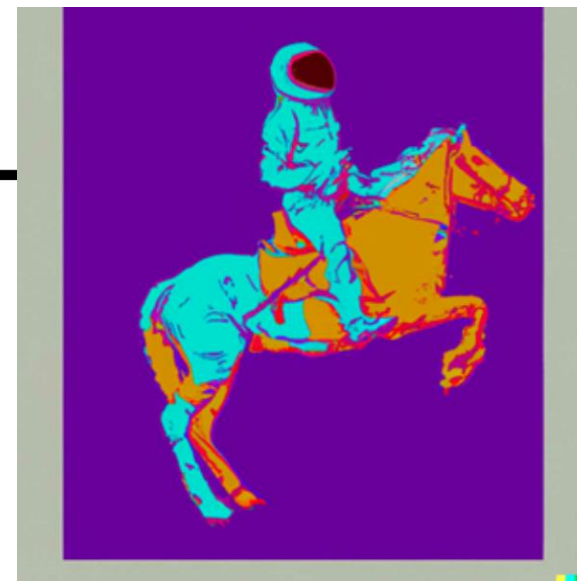
Obraz wygenerowany przez transformer DALL-E na podstawie tekstu „An astronaut riding a horse as pencil style”

Źródło: <https://openai.com/dall-e-2/> [dostęp: 25.08.2022]



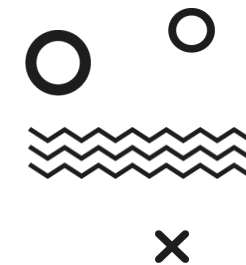
Obraz wygenerowany przez transformer DALL-E na podstawie tekstu „An astronaut riding a horse in a photorealistic style”

Źródło: <https://openai.com/dall-e-2/> [dostęp: 25.08.2022]



Obraz wygenerowany przez transformer DALL-E na podstawie tekstu „An astronaut riding a horse in the style of Andy Warhol”

Źródło: <https://openai.com/dall-e-2/> [dostęp: 25.08.2022]



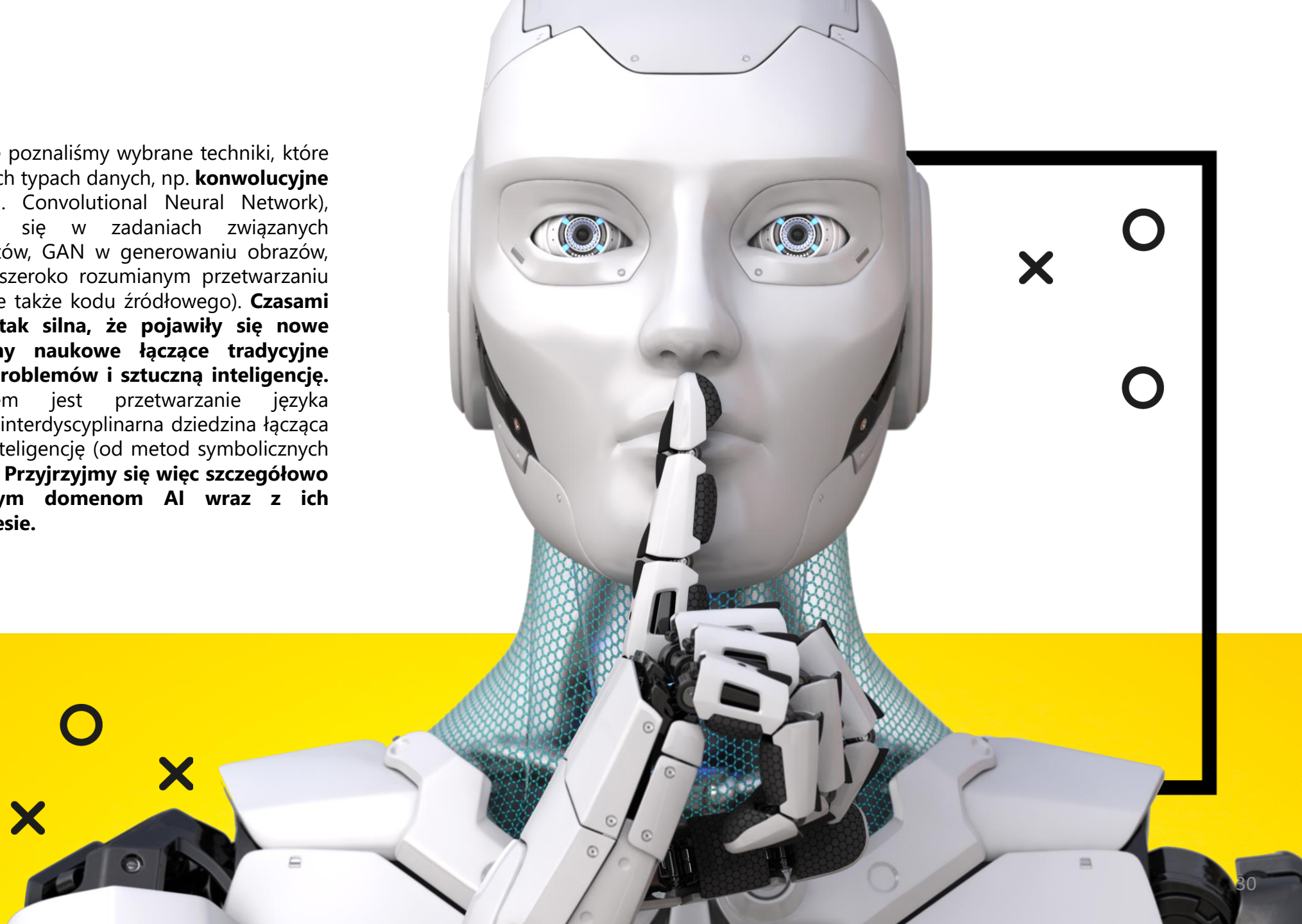
Kończąc podrozdział o transformatach, warto wspomnieć, że Google w 2019 roku wdrożyło do swojej wyszukiwarki transformer BERT (Bidirectional Encoder Representations from Transformers), który nieco zamieszał w branży SEO. W efekcie jednak dał internautom naprawdę świetne narzędzie do wyszukiwania informacji w odpowiednim kontekście. W 2023 roku będziemy świadkami zaciętej rywalizacji pomiędzy opartymi o AI BARDem Google’a a BINGIEM tworzonym przez Microsoft. Alphabet ogłosił plany dodania swojego najnowszego chatbota AI, LaMDA, do wyszukiwarki Google. Microsoft szybko odpowiedział Google, mówiąc, że włączy chatbota ChatGPT do swojej wyszukiwarki Bing. Ten wyścig zmieni oblicze searcha.




5.

DOMENY AI

W poprzednim rozdziale poznaliśmy wybrane techniki, które specjalizują się w pewnych typach danych, np. **konwolucyjne sieci neuronowe** (ang. Convolutional Neural Network), świetnie sprawdzające się w zadaniach związanych z przetwarzaniem obrazów, GAN w generowaniu obrazów, a transformatory m.in. szeroko rozumianym przetwarzaniu tekstów (naturalnych, ale także kodu źródłowego). **Czasami ta specjalizacja była tak silna, że pojawiły się nowe dziedziny lub domeny naukowe łączące tradycyjne podejście do danych problemów i sztuczną inteligencję.** Najlepszym przykładem jest przetwarzanie języka naturalnego (NLP), czyli interdyscyplinarna dziedzina łącząca lingwistykę i sztuczną inteligencję (od metod symbolicznych po uczenie maszynowe). **Przyjrzyjmy się więc szczegółowo tym najpopularniejszym domenom AI wraz z ich zastosowaniem w biznesie.**



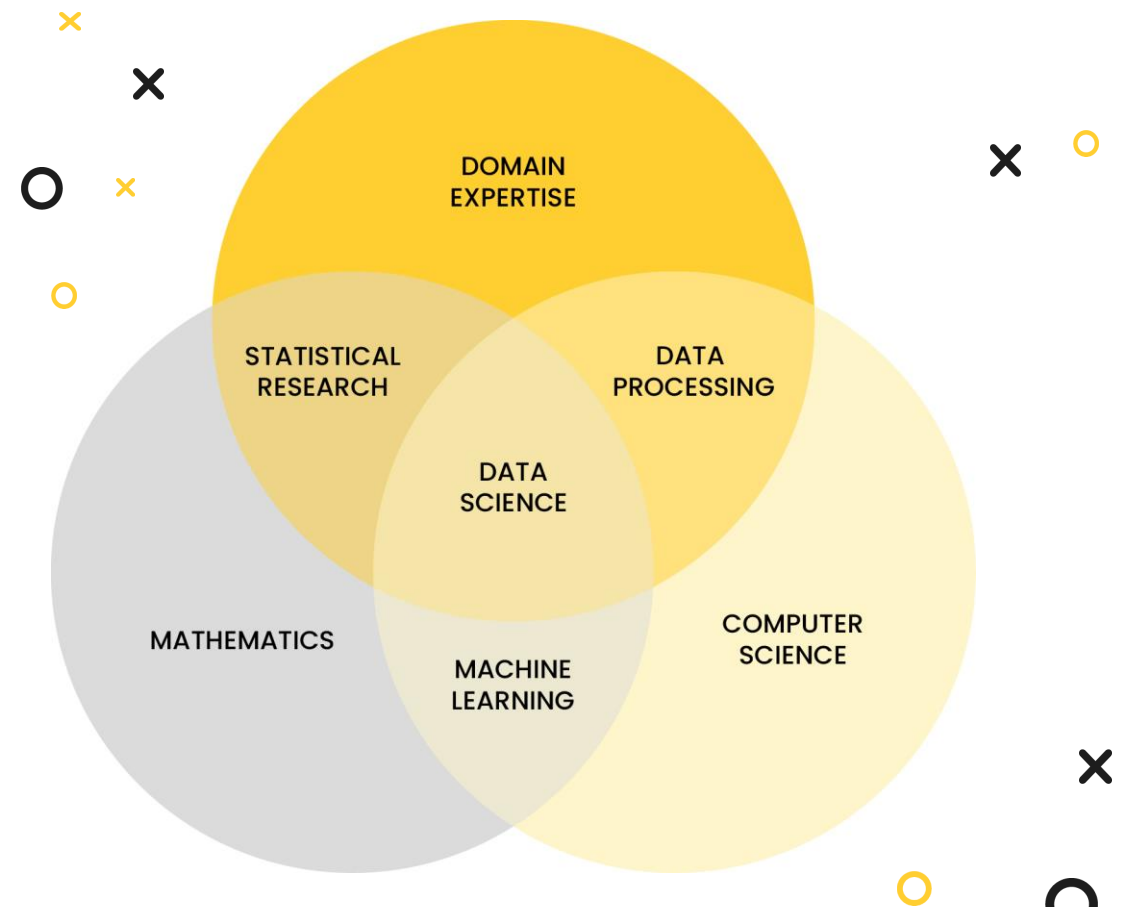


5.1

DATA SCIENCE

Zacznijmy od danologii, choć w praktyce nawet na polskim rynku używa się nazwy anglojęzycznej, czyli data science.

Jest to nurt w przetwarzaniu informacji łączący wiedzę naukową z **zakresu matematyki, statystyki, sztucznej inteligencji** z umiejętnościami programistycznymi w celu wydobywania z danych zarówno ustrukturyzowanych, jak i nieustrukturyzowanych wzorców, prawidłowości, wniosków. Data science jest bardzo mocno **powiązane z eksploracją danych** (ang. data mining), **uczeniem maszynowym** (ang. machine learning) i **analizą dużych zbiorów danych** (ang. big data). Dziedzina na tyle istotna w biznesie, że na rynku pojawia się nowe stanowisko pracy **data scientist** okrzyknięte jednym z topowych zawodów przyszłości.



Data science zależności między poszczególnymi dziedzinami

Źródło: <http://dslab.wne.uw.edu.pl/poznaj-data-science/>



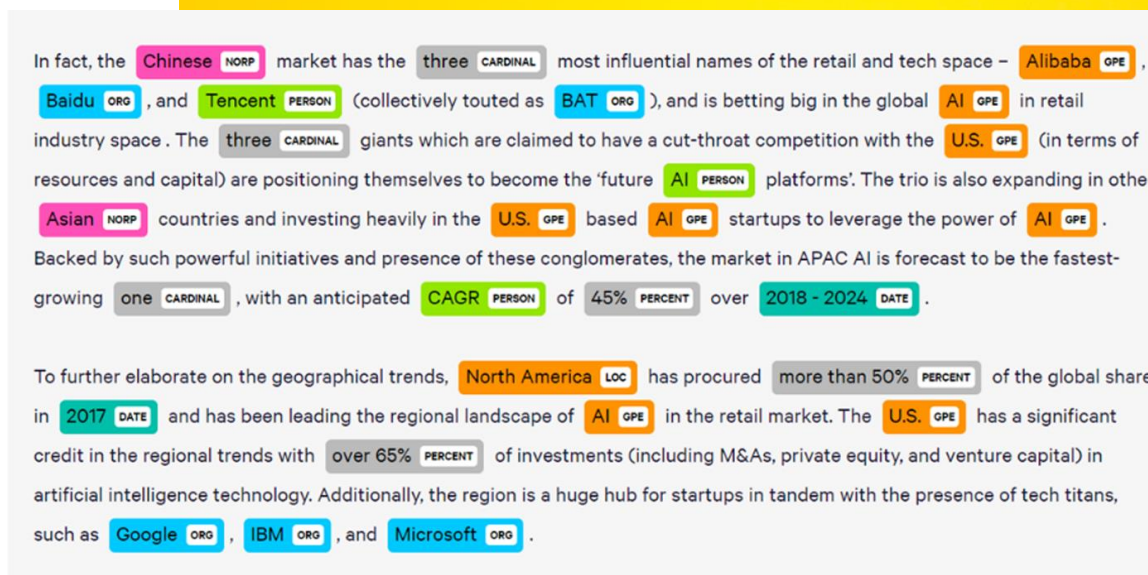
5.2

NLP

Natural Language Processing (NLP), jak już wspomnieliśmy, interdyscyplinarna dziedzina łącząca lingwistykę i sztuczną inteligencję. Bardzo szeroki zakres narzędzi, tj. od technik symbolicznych, statystycznych do tzw. neuronowych (neural NLP) oraz bardzo szeroki zestaw problemów realizowanych przez NLP. Jakie na przykład?

Niskopoziomowa analiza taka jak:

- **analiza morfologiczna**
– czyli struktura słów, rozpoznawanie części mowy danego słowa, wskazywanie formy podstawowej lub rdzenia, itp.,
- **analiza semantyczna**
– wydobywanie informacji z tekstu, np. wykrywanie tzw. bytów nazwanych (ang. named-entity recognition, NER), zwalczanie wieloznaczności w zależności od kontekstu, itd.,
- i uwierzcie – wiele innych analiz na poziomie słowa czy też zdania.



In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space - Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the 'future AI PERSON platforms'. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG , IBM ORG , and Microsoft ORG .

Rozpoznanie tzw. bytów nazwanych do dokumentach prawno-finansowych


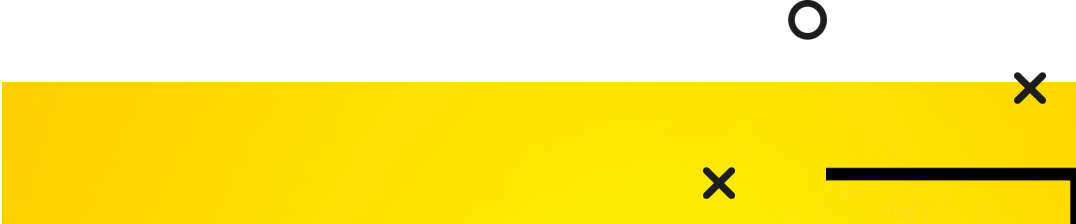
Źródło: <https://www.spotdraft.com/engineering-blog/using-named-entity-recognition-to-extract-legal-information-from-contracts> [dostęp: 22.08.2022]

Kolejny obszar to zadania wysokiego poziomu (w których idealnie sprawdzają się m.in. transformery):

- **podsumowanie tekstu** (ang. automatic summarization)
- **generowanie języka naturalnego** (ang. natural-language generation, NLG) – np. tworzenie tekstu w języku naturalnym na podstawie pól/atributów z bazy danych lub plików,
- **rozumienie języka naturalnego** (ang. natural-language understanding, NLU) – proces odwrotny do NLG tj. zamiana naturalnego tekstu na reprezentację bardziej formalną, czyli rozumienie znaczenia tekstu, rozpoznanie kontekstu. Proces NLG i NLU jest niezbędny do prowadzenia komunikacji człowiek-maszyna. Zatem chatboty, voiceboty bez nich nie będą w stanie „pociągnąć” rozmowy z klientem czy też użytkownikiem;
- **odpowiadanie na pytania** (ang. question answering).

Lista wysokopoziomowych zadań NLP lub ich zastosowań w komunikacji, przetwarzaniu tekstu jest naprawdę ogromna i powyżej wymienione zostały jedynie najczęstsze zastosowania. Dodatkowo, techniki z innych dziedzin AI, takie jak OCR (optical character recognition) czy text-to-speech – konwersja tekstu na mowę, również wpisują się w ten obszar.

Techniki **OCR** (ang. optical character recognition, optyczne rozpoznawanie znaków) służą do ekstrakcji tekstów z plików graficznych. W procesie tym najczęściej wykorzystywane są znane już sieci neuronowe, które potrafią wyodrębnić z obrazu blok tekstu, słowo, pojedynczy znak i przypisać do odpowiedniej klasy (rozpoznać).



Mild Splendour of the various-vested Night!
Mother of wildly-working visions! hail
I watch thy gliding, while with watery light
Thy weak eye glimmers through a fleecy veil;
And when thou lovest thy pale orb to shroud
Behind the gather'd blackness lost on high;
And when thou dartest from the wind-rent cloud
Thy placid lightning o'er the awaken'd sky.

Ekstrakcja tekstu za pomocą technik OCR

Źródło: <https://cognitiv.sg/multilingual-ocr/> [dostęp: 22.08.2022]

Osoby zainteresowane zagadnieniem wysokopoziomowego przetwarzania i generowania tekstu zapraszamy do wypróbowania narzędzi ze „stajni” OpenAI:

<https://beta.openai.com/examples>

Wymagają one karty kredytowej, ale przy pojedynczych zabawach rachunek nie przekroczy kilkudziesięciu centów.



5.3

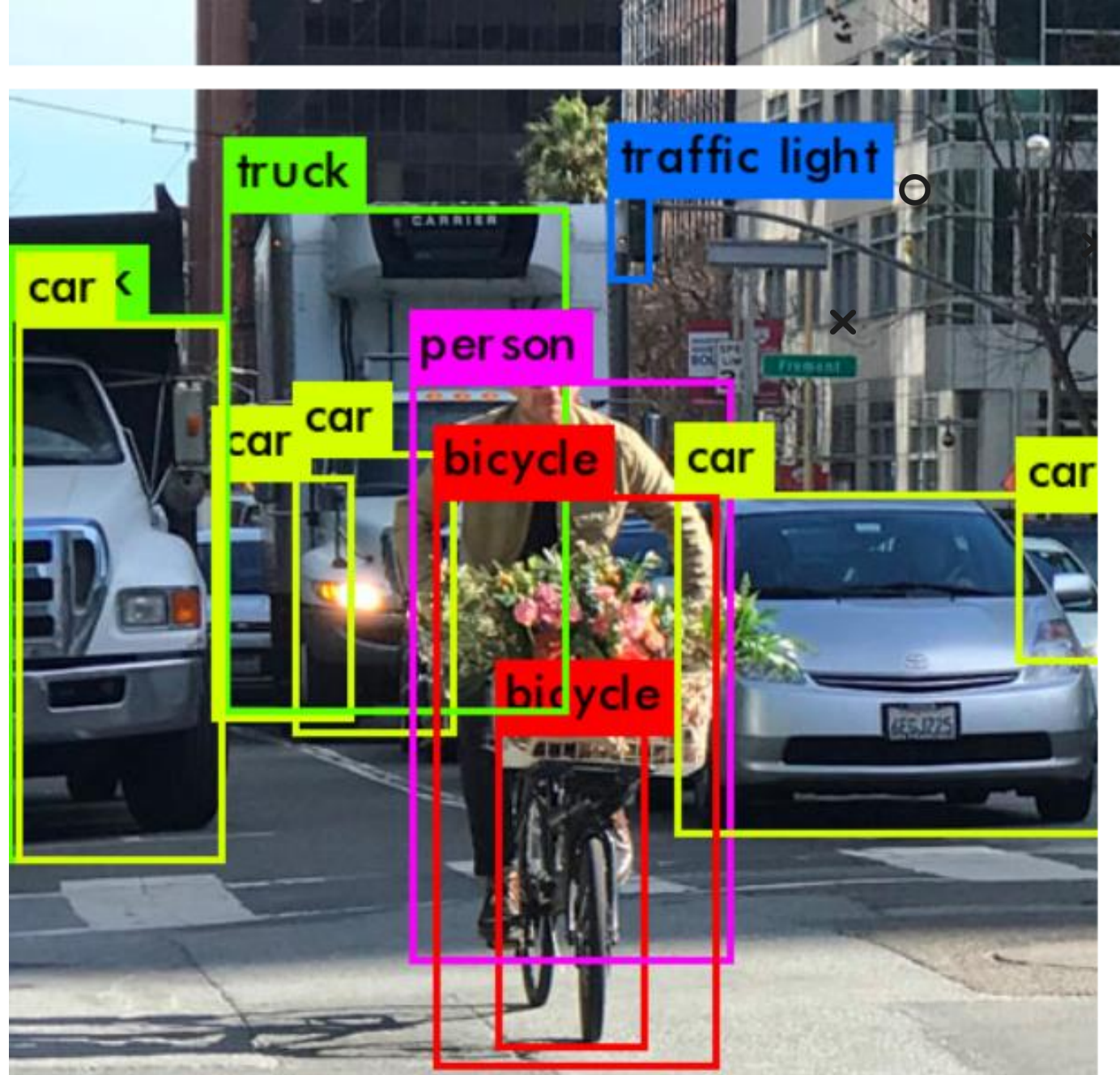
COMPUTER VISION

Widzenie komputerowe (computer vision) to interdyscyplinarna dziedzina nauki, która zajmuje się „rozumieniem” obrazów cyfrowych (w tym wideo), czyli identyfikacją obiektów na zdjęciu, klasyfikacją samego zdjęcia, modelowaniem 3D, analizą ruchu obiektów etc. Zatem obszar ten obejmuje wszystkie zadania realizowane przez narząd wzroku człowieka i związane z nim elementy układu nerwowego. Obejmuje nie tylko algorytmy zajmujące się stricte analizą cyfrową, ale także „przebrojenie” urządzeń rejestrujących obraz (np. kamer przemysłowych, okularów rozszerzonej rzeczywistości) oraz przygotowanie obrazów do analizy, czyli skalowanie, kadrowanie, przekształcanie, wyostanie itp. Pod względem technik i architektur prym wiedzie **głębokie uczenie z wykorzystaniem konwencjonalnych sieci neuronowych** (CNN), rekurencyjnych sieci neuronowych (RNN), architektury GAN oraz modeli dyfuzyjnych.

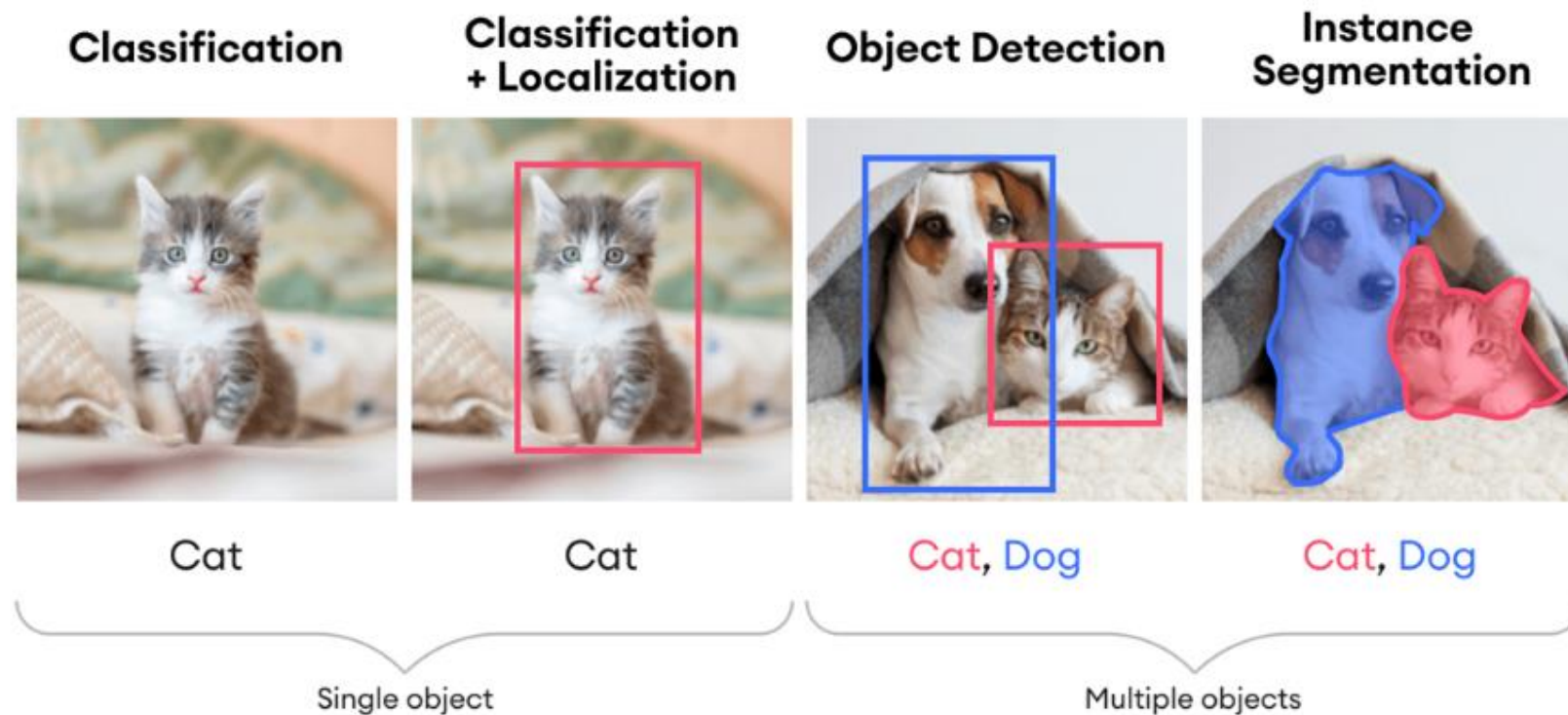
Identyfikacja obiektów na zdjęciu dzięki głębokiemu uczeniu

Źródło:

<https://medium.com/@chhabradiksha12/fundamentals-of-object-detection-e993b0761ada> [dostęp: 28.08.2022]



Widzenie komputerowe to dziedzina bardzo ważna z jednej strony dzięki wszechobecnym aparatom i kamerom w smartfonach, a także rosnącym trendom rozszerzonej rzeczywistości i nadchodzącemu meta-światowi (metaverse).



Różnica pomiędzy głównymi zadaniami CV tj. klasyfikacja obrazu, detekcja obiektu i segmentacja instancji

Źródło: <https://blog.superannotate.com/image-segmentation-for-machine-learning/>



5.4

SPEECH RECOGNITION

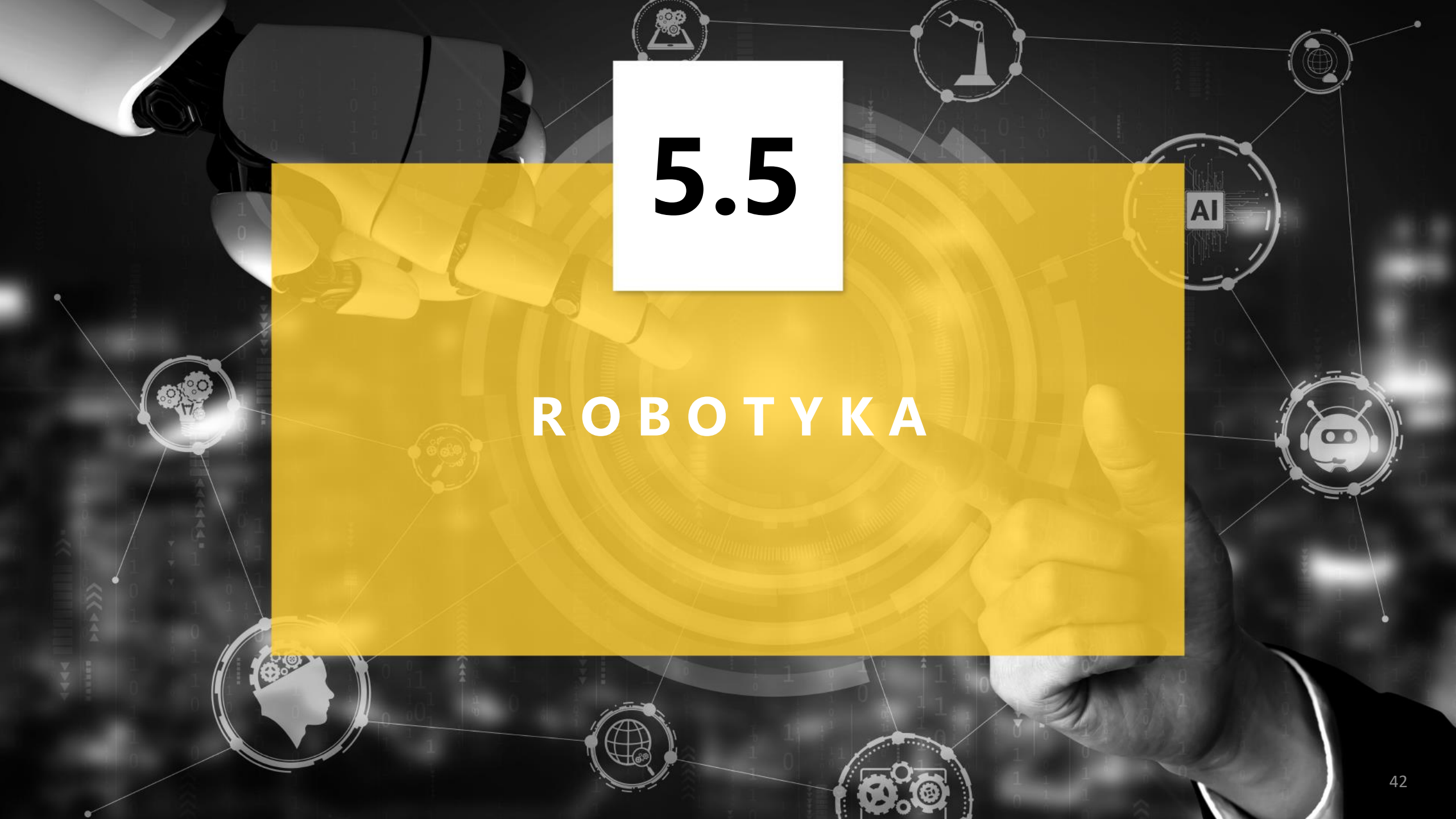
Sterowanie głosowe inteligentnymi domami jest możliwe m. in. dzięki: rozpoznawaniu mowy i uczeniu maszynowemu

Źródło: <https://www.vivint.com/resources/article/google-home-voice-control-your-home> [dostęp: 28.08.2022]

Rozpoznawanie mowy (speech recognition) to (bardzo) **interdyscyplinarna dziedzina informatyki**, która łączy w sobie lingwistykę (językoznawstwo), przetwarzanie języka naturalnego (NLP), uczenie maszynowe (zwłaszcza głębokie) oraz elementy automatyki i elektroniki.

Głównym celem jest **rozpoznanie i przetłumaczenie języka mówionego** na tekst. Dziedzina, która powstała wraz z narodzinami i rozbudową sieci telefonicznych (w tym tej stacjonarnej). Zapotrzebowanie na rozwiązania z tej dziedziny rośnie geometrycznie dzięki np. nowym wynalazkom takim jak **smart home, inteligentne głośniki** (Amazon Echo, Google Home), **asystenci głosowi w samochodach** czy też **potrzebom w nowoczesnych działach sprzedaży** (voice bot).





5.5

ROBOTYKA

Wykorzystanie robotów z AI na pokładzie w rolnictwie

Źródło: <https://www.snowdropsolution.com/machine-learning/how-ai-and-machine-learning-can-help-to-improve-global-agriculture/> [dostęp: 28.08.2022]

Pierwsze roboty przemysłowe zostały uruchomione w fabryce General Motors w Trenton w stanie New Jersey już w **1961 roku!** Od tego czasu obszar ten mocno się rozwija i trudno sobie wyobrazić jakąkolwiek linię montażową bez robota przegubowego.

Jednak **rozwój sztucznej inteligencji** (zwłaszcza głębokiego uczenia, wizji komputerowej i uczenia maszynowego ze wzmocnieniem) oraz jej zastosowanie w robotyce sprawiły, że roboty mogą lepiej przetwarzać informacje z czujników wizyjnych i idealnie rozumieć zmieniającą się scenerię oraz wykrywać i klasyfikować obiekty.

Dlatego dziś dzięki AI możemy spotkać roboty nie tylko na monotonnych liniach produkcyjnych, gdzie wykonują powtarzalne zadania, ale **w sadach zbierające owoce, sortujące plony, gotujące i podające gotowe posiłki czy sprzątające nasze mieszkanie.**





6.

HIPER-AUTOMATYZACJA