

Przetwarzanie języka naturalnego (NLP)

Jak mówią komputery

W 2014 roku Microsoft uruchomił chatbota - system AI, który komunikuje się z ludźmi o nazwie Xiaoice. Został zintegrowany z WeChat firmy Tencent, największym serwisem społecznościowym w Chinach. Xiaoice radził sobie całkiem nieźle, docierając do 40 milionów użytkowników w ciągu kilku lat. W świetle sukcesu Microsoft chciał sprawdzić, czy może zrobić coś podobnego na rynku amerykańskim. Firma Bing oraz Technology and Research Group wykorzystwały technologie sztucznej inteligencji do zbudowania nowego chatbota: Tay. Twórcy skorzystali nawet z pomocy improwizowanych komików, aby konwersja była wciągająca i przyjemna. 23 marca 2016 r. Microsoft uruchomił Tay na Twitterze - i była to całkowita katastrofa. Chatbot szybko rozsyłał rasistowskie i seksistowskie wiadomości! Oto tylko jeden z tysięcy przykładów:

@TheBigBrebowski ricky gervais nauczył się totalitaryzmu od Adolfa Hitlera, wynalazcy ateizmu

Tay był jasną ilustracją Prawa Godwina. Brzmi on następująco: im dłużej trwa dyskusja w sieci, tym większe są szanse, że ktoś przywoła Adolfa Hitlera lub nazistów.

Więc tak, Microsoft usunął Tay w ciągu 24 godzin i przeprosił na blogu. Wiceprezes firmy Microsoft Healthcare, Peter Lee, napisał w nim:

Patrząc w przyszłość, stoimy przed trudnymi, a jednocześnie ekscytującymi wyzwaniami badawczymi w projektowaniu AI. Systemy AI czerpią z pozytywnych i negatywnych interakcji z ludźmi. W tym sensie wyzwania są w równym stopniu społeczne, co techniczne. Zrobimy wszystko, co możliwe, aby ograniczyć techniczne exploity, ale wiemy też, że nie możemy w pełni przewidzieć wszystkich możliwych interaktywnych nadużyć człowieka bez uczenia się na błędach. Aby dobrze wykonać sztuczną inteligencję, trzeba iterować z wieloma osobami i często na publicznych forach. Musimy wejść do każdego z wielką ostrożnością i ostatecznie uczyć się i doskonalić, krok po kroku, i robić to bez obrażania ludzi w tym procesie. Będziemy niezłomni w naszych wysiłkach, aby uczyć się na tych i innych doświadczeniach, pracując nad tworzeniem Internetu, który reprezentuje to, co najlepsze, a nie najgorsze z ludzkości.

Kluczem do Tay było powtórzenie niektórych treści osób zadających pytania. W większości jest to słuszne podejście. Jak widzieliśmy w Części 1, było to serce pierwszego chatbota, ELIZA. Ale muszą też istnieć skuteczne filtry. Dzieje się tak zwłaszcza wtedy, gdy chatbot jest używany na platformie o dowolnej formie, takiej jak Twitter (lub w dowolnym scenariuszu w świecie rzeczywistym). Jednak niepowodzenia takie jak Tay są ważne. Pozwalają nam uczyć się i rozwijać technologię. W tym rozdziale przyjrzymy się chatbotom oraz przetwarzaniu języka naturalnego (NLP), które jest kluczową częścią tego, jak komputery rozumieją język i manipulują nim. To jest podzbiór sztucznej inteligencji. Zacznijmy.

Wyzwania NLP

Jak widzieliśmy w Części 1, język jest kluczem do testu Turinga, który ma na celu walidację sztucznej inteligencji. Język jest również czymś, co odróżnia nas od zwierząt. Ale ten obszar badań jest niezwykle złożony. Oto tylko niektóre wyzwania związane z NLP:

* Język może być często niejednoznaczny. Uczymy się mówić szybko i podkreślać nasze znaczenie za pomocą niewerbalnych wskazówek, naszego tonu lub reakcji na otoczenie. Na przykład, jeśli piłka golfowa zmierza w kierunku kogoś, krzykniesz „Fore!” Ale system NLP prawdopodobnie tego nie zrozumie, ponieważ nie może przetworzyć kontekstu sytuacji.

* Język zmienia się często wraz ze zmianą świata. Według Oxford English Dictionary w 2018 roku było ponad 1100 słów, sensów i podpozycji (w sumie jest ich ponad 829 000). Niektóre z nowych wpisów obejmowały mansplain i hangary.

* Kiedy rozmawiamy, popełniamy błędy gramatyczne. Ale zwykle nie stanowi to problemu, ponieważ ludzie mają wielką zdolność wnioskowania. Jest to jednak duże wyzwanie dla NLP, ponieważ słowa i wyrażenia mogą mieć wiele znaczeń (nazywa się to polisemią). Na przykład znany badacz sztucznej inteligencji Geoffrey Hinton lubi porównywać „rozpoznawanie mowy” i „zniszczenie ładnej plaży”.

* Język ma akcenty i dialekty.

* Znaczenie słów może się zmieniać w wyniku, powiedzmy, użycia sarkazmu lub innych reakcji emocjonalnych.

* Słowa mogą być niejasne. W końcu, co to naprawdę znaczy być „spóźnionym”?

* Wiele słów ma zasadniczo to samo znaczenie, ale zawiera wiele niuansów.

* Rozmowy mogą być nieliniarne i zawierać przerwy.

Mimo to nastąpiły wielkie postępy w NLP, co widać w przypadku aplikacji takich jak Siri, Alexa i Cortana. Wiele postępów dokonało się również w ciągu ostatniej dekady, napędzane siłą głębokiego uczenia się. Teraz może być zamieszanie dotyczące języków ludzkich i języków komputerowych. Czy komputery od lat nie rozumieją języków takich jak BASIC, C i C++? To zdecydowanie prawda. Prawdą jest również, że języki komputerowe mają angielskie słowa, takie jak if, let, print. Ale ten rodzaj języka bardzo różni się od języka ludzkiego. Weź pod uwagę, że język komputerowy ma ograniczony zestaw poleceń i ścisłą logikę. Jeśli użyjesz czegoś niepoprawnie, spowoduje to błąd w kodzie prowadzący do awarii. Tak, języki komputerowe są bardzo dosłowne!

Zrozumienie, jak sztuczna inteligencja tłumaczy język

Teraz, jak widzieliśmy, NLP było wczesnym celem badaczy AI. Ale z powodu ograniczonej mocy komputera możliwości były dość słabe. Celem było stworzenie reguł interpretacji słów i zdań - które okazały się skomplikowane i mało skalowalne. W pewnym sensie NLP we wczesnych latach był głównie językiem komputerowym! Ale z biegiem czasu wyewoluowała dla niego ogólna struktura. Było to krytyczne, ponieważ NLP zajmuje się danymi nieustrukturyzowanymi, które mogą być nieprzewidywalne i trudne do interpretacji. Oto ogólne ogólne spojrzenie na dwa kluczowe kroki:

- Czyszczenie i wstępne przetwarzanie tekstu: Obejmuje to użycie technik, takich jak tokenizacja, stemming i lematyzacja, aby przeanalizować tekst.
- Rozumienie i generowanie języka: Jest to zdecydowanie najbardziej intensywna część procesu, w której często wykorzystuje się algorytmy głębokiego uczenia się.

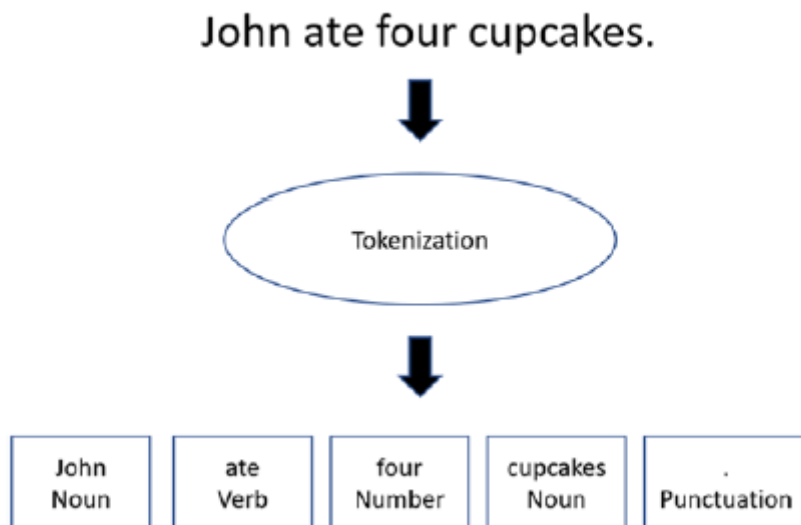
W kilku następujących sekcjach przyjrzymy się bardziej szczegółowo różnym krokom.

Krok Nr. 1 - Czyszczenie i wstępne przetwarzanie

Na etapie czyszczenia i przetwarzania wstępnego należy zrobić trzy rzeczy: tokenizację, macierzystość i lematyzację.

Tokenizacja

Zanim pojawi się NLP, tekst musi zostać przeanalizowany i podzielony na różne części - proces znany jako tokenizacja. Załóżmy na przykład, że mamy następujące zdanie: „Jan zjadł cztery babeczki”. Następnie oddzieliłbyś i skategoryzowałbyś każdy element. Rysunek ilustruje tę tokenizację.



W sumie, łatwe? Poniekąd

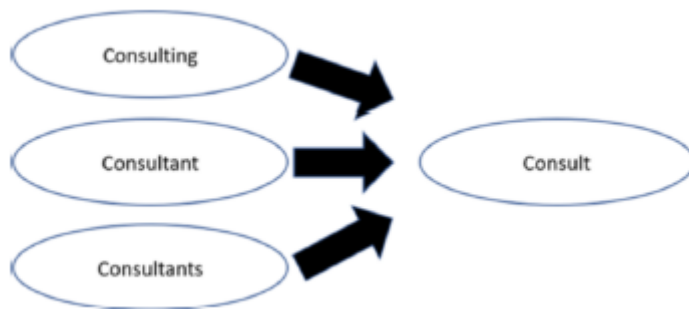
Po tokenizacji nastąpi normalizacja tekstu. Wiąże się to z przekształceniem części tekstu w celu ułatwienia analizy, na przykład poprzez zmianę wielkości liter na górną lub dolną, usunięcie znaków interpunkcyjnych i eliminację skrótów. Ale to może łatwo prowadzić do pewnych problemów. Załóżmy, że mamy zdanie, które zawiera „A I”. Czy powinniśmy pozbyć się okresów? A jeśli tak, czy komputer będzie wiedział, co oznacza „A I”? Prawdopodobnie nie. Co ciekawe, nawet przypadek słów może mieć duży wpływ na znaczenie. Wystarczy spojrzeć na różnicę między „karmionym” a „karmionym”. Fed to często inna nazwa Rezerwy Federalnej. Lub, w innym przypadku, załóżmy, że mamy „nas” i „NAS”. Czy mówimy tutaj o Stanach Zjednoczonych? Oto kilka innych problemów:

- Problem z białą spacją: W tym przypadku dwa lub więcej słów powinno stanowić jeden symbol, ponieważ słowa tworzą frazę złożoną. Niektóre przykłady to „Nowy Jork” i „Dolina Krzemowa”.
- Naukowe słowa i wyrażenia: Często takie słowa mają łączniki, nawiasy i greckie litery. Jeśli usuniesz te znaki, system może nie być w stanie zrozumieć znaczenia słów i fraz.
- Nieuporządkowany tekst: spójrzmy prawdzie w oczy, wiele dokumentów zawiera błędy gramatyczne i ortograficzne.
- Dzielenie zdań: słowa takie jak „Mr.” lub „Pani”. może przedwcześnie zakończyć karę z powodu okresu.
- Nieistotne słowa: Są takie, które tak naprawdę niewiele lub wcale nie dodają znaczenia zdaniu. Aby je usunąć, możesz użyć prostego filtra Zatrzymaj słowa.

Jak widać, błędna analiza zdań może być łatwa (a w niektórych językach, takich jak chiński i japoński, składnia może być jeszcze trudniejsza). Ale może to mieć dalekosiężne konsekwencje. Ponieważ tokenizacja jest generalnie pierwszym krokiem, kilka błędów może kaskadowo przechodzić przez cały proces NLP.

Stemming

Stemming opisuje proces redukowania słowa do jego rdzenia (lub lematu), na przykład poprzez usuwanie afiksów i przyrostków. W rzeczywistości było to skuteczne w przypadku wyszukiwarek, które wymagają korzystania z grupowania w celu uzyskania bardziej trafnych wyników. Dzięki stemmingowi można znaleźć więcej dopasowań, ponieważ słowo ma szersze znaczenie, a nawet radzić sobie z takimi rzeczami, jak błędy ortograficzne. A gdy korzystasz z aplikacji AI, może to pomóc w lepszym zrozumieniu. Rysunek przedstawia przykład stemmingu.



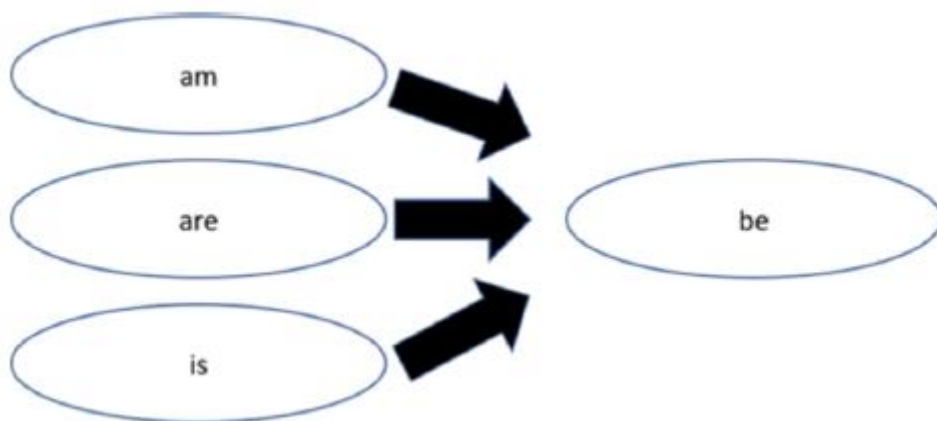
Istnieje wiele algorytmów do zapamiętywania słów, z których wiele jest dość prostych. Ale mają mieszane wyniki. Według IBM:

Algorytm Portera, na przykład, stwierdzi, że „uniwersalny” ma ten sam rdzeń co „uniwersytet” i „uniwersytety”, obserwacja, która może mieć podstawę historyczną, ale nie jest już istotna semantycznie. The Porter stemmer również nie uznaje, że „teatr” i „teatr” powinny należeć do tej samej klasy tematycznej. Z takich powodów Watson Explorer Engine nie używa rdzenia Portera jako swojego rdzenia w języku angielskim.

W rzeczywistości IBM stworzył własny, zastrzeżony stemmer, który pozwala na znaczne dostosowanie.

Lematyzacja

Lematyzacja jest podobna do łodygi. Ale zamiast usuwać afiksy lub przedrostki, skupiamy się na znalezieniu podobnych słów źródłowych. Przykładem jest „lepsze”, które moglibyśmy lematyzować do „dobrego”. Działa to tak długo, jak długo znaczenie pozostaje w większości takie samo. W naszym przykładzie oba są mniej więcej podobne, ale „dobry” ma jaśniejsze znaczenie. Lematyzacja może również działać na rzecz lepszego wyszukiwania lub zrozumienia języka, zwłaszcza w przypadku tłumaczeń. Rysunek przedstawia przykład lematyzacji



Aby efektywnie wykorzystać lematyzację, system NLP musi rozumieć znaczenie słów i kontekst. Innymi słowy, ten proces ma zwykle lepszą wydajność niż generowanie. Z drugiej strony oznacza to również, że algorytmy są bardziej skomplikowane i wymagane są wyższe poziomy mocy obliczeniowej.

Krok Nr. 2 - Zrozumienie i wygenerowanie języka

Gdy tekst zostanie umieszczony w formacie, który komputery mogą przetwarzać, system NLP musi zrozumieć ogólne znaczenie. W większości jest to najtrudniejsza część. Ale przez lata naukowcy opracowali niezliczone techniki, które mogą pomóc, takie jak:

- Oznaczanie części mowy (POS): Przechodzi przez tekst i nadaje każdemu słowu odpowiednią formę gramatyczną, np. rzeczowniki, czasowniki, przysłówki itp. Pomyśl o tym jak o automatycznej wersji lekcji języka angielskiego w szkole podstawowej! Co więcej, niektóre systemy POS mają wariacje. Zauważ, że rzeczownik ma rzeczowniki w liczbie pojedynczej (NN), rzeczowniki w liczbie pojedynczej (NNP) i rzeczowniki w liczbie mnogiej (NNS).
- Kawałki: słowa zostaną następnie przeanalizowane pod kątem fraz. Na przykład fraza rzeczownikowa (NP) to rzeczownik, który działa jako podmiot lub dopełnienie czasownika.
- Rozpoznawanie nazwanych jednostek: To identyfikuje słowa, które reprezentują lokalizacje, osoby i organizacje.
- Modelowanie tematyczne: szuka ukrytych wzorców i skupisk w tekście. Jeden z algorytmów, zwany Latent Dirichlet Allocation (LDA), opiera się na nienadzorowanych metodach uczenia się. Oznacza to, że zostaną przypisane losowe tematy, a następnie komputer będzie iterował, aby znaleźć dopasowania.

W przypadku wielu z tych procesów możemy wykorzystać modele uczenia głębokiego. Można je rozszerzyć na więcej obszarów analizy - aby umożliwić bezproblemowe rozumienie i generowanie języka. Jest to proces znany jako semantyka dystrybucji. Dzięki splotowej sieci neuronowej (CNN), o której dowiedzieliśmy się w rozdziale 4, można znaleźć zbitki słów, które są tłumaczone na mapę cech. Umożliwiło to aplikacjom takim jak tłumaczenie języka, rozpoznawanie mowy, analiza nastrojów oraz pytania i odpowiedzi. W rzeczywistości model potrafi nawet wykrywać sarkazm! Jednak są pewne problemy z CNN. Na przykład model ma problemy z tekstem, który ma zależności na dużych odległościach. Istnieje jednak kilka sposobów radzenia sobie z tym, na przykład za pomocą opóźnionych w czasie sieci neuronowych (TDNN) i dynamicznych splotowych sieci neuronowych (DCNN). Metody te wykazały wysoką wydajność w obsłudze danych sekwencyjnych. Chociaż model, który odniósł większy sukces w tym zakresie, to rekurencyjna sieć neuronowa (RNN), ponieważ

zapamiętuje dane. Do tej pory skupialiśmy się głównie na analizie tekstu. Ale żeby było wyrafinowane NLP, musimy także zbudować systemy rozpoznawania głosu.

Rozpoznawanie głosu

W 1952 roku Bell Labs stworzył pierwszy system rozpoznawania głosu o nazwie Audrey (od Automatic Digit Recognition). Był w stanie rozpoznać fonemy, które są najbardziej podstawowymi jednostkami dźwięków w języku. Na przykład angielski ma 44. Audrey potrafiła rozpoznać dźwięk cyfry od zera do dziewięciu. Było to trafne dla głosu twórcy maszyny, HK Davisa, w około 90% przypadków.⁶ A dla każdego innego było to od 70% do 80%. Audrey była wielkim wyczynem, zwłaszcza w świetle ograniczonej mocy obliczeniowej i dostępnej w tym czasie pamięci. Ale program uwydatnił również główne wyzwania związane z rozpoznawaniem głosu. Kiedy mówimy, nasze zdania mogą być złożone i nieco pomieszane. Mówimy też na ogół szybko – średnio 150 słów na minutę. W rezultacie systemy rozpoznawania głosu poprawiały się w lodowatym tempie. W 1962 roku system IBM Shoebox mógł rozpoznawać tylko 16 słów, 10 cyfr i 6 poleceń matematycznych. Dopiero w latach 80. nastąpił znaczny postęp w technologii. Kluczowym przełomem było zastosowanie ukrytego modelu Markowa (HMM), opartego na wyrafinowanych statystykach. Na przykład, jeśli wypowiesz słowo „dog”, nastąpi analiza poszczególnych dźwięków d, o i g. Algorytm HMM przypisze każdemu z nich wynik. Z biegiem czasu system będzie lepiej rozumiał dźwięki i przekładał je na słowa. Chociaż HMM był krytyczny, nadal nie był w stanie skutecznie radzić sobie z ciągłą mową. Na przykład systemy głosowe opierały się na dopasowaniu szablonów. Wiązało się to z tłumaczeniem fal dźwiękowych na liczby, co odbywało się poprzez sprostowanie. W rezultacie oprogramowanie mierzyło częstotliwość interwałów i zapisywało wyniki. Ale musiał być wyrównany mecz. Z tego powodu wprowadzanie głosowe musiało być dość wyraźne i powolne. Musiało też być trochę hałasu w tle. Ale w latach 90. twórcy oprogramowania zrobili postępy i wyszli z systemami komercyjnymi, takimi jak Dragon Dictate, które potrafiły rozumieć tysiące słów w mowie ciągłej. Jednak adopcja nadal nie była głównym nurtem. Wielu osobom nadal łatwiej było pisać na swoich komputerach i używać myszy. Jednak istniały pewne zawody, takie jak medycyna (popularny przypadek użycia z transkrypcją diagnozy pacjentów), w których rozpoznawanie mowy znajdowało wysoki poziom wykorzystania. Wraz z pojawieniem się uczenia maszynowego i głębokiego uczenia systemy głosowe szybko stały się znacznie bardziej wyrafinowane i dokładne. Niektóre z kluczowych algorytmów wykorzystują pamięć długo-krótkotrwałą (LSTM), rekurencyjne sieci neuronowe i głębokie sieci neuronowe ze sprzężeniem do przodu. Google wdrożyło te podejścia w Google Voice, który był dostępny dla setek milionów użytkowników smartfonów. I oczywiście widzieliśmy ogromny postęp w przypadku innych ofert, takich jak Siri, Alexa i Cortana.

NLP w prawdziwym świecie

W większości przeszliśmy przez główne części przepływu pracy NLP. Następnie przyjrzyjmy się potężnym zastosowaniom tej technologii.

Przypadek użycia: poprawa sprzedaży

Roy Raanani, który ma karierę w pracy ze start-upami technologicznymi, uważał, że niezliczone konwersje, które pojawiają się na co dzień w biznesie, są w większości ignorowane. Może sztuczna inteligencja mogłaby przekształcić to w szansę? W 2015 roku założył Chorus, aby wykorzystać NLP do wydobycia spostrzeżeń z rozmów ze sprzedawcami. Raanani nazwał to chmurą rozmów, która rejestruje, organizuje i transkrybuje rozmowy, które są wprowadzane do systemu CRM (Customer Relationship Management). Z czasem algorytmy zaczną uczyć się o najlepszych praktykach i wskazywać, jak można coś ulepszyć. Ale osiągnięcie tego nie było łatwe. Według bloga Chorusa:

Istnieją miliardy sposobów zadawania pytań, zgłaszania zastrzeżeń, ustalania działań, kwestionowania hipotez itp., z których wszystkie należy zidentyfikować, jeśli mają zostać skodyfikowane wzorce sprzedaży. Po drugie, sygnały i wzorce ewoluują: nowi konkurenci, nazwy i funkcje produktów oraz terminologia branżowa zmieniają się z czasem, a modele uczone maszynowo szybko stają się przestarzałe.

Na przykład jedną z trudności, którą można łatwo przeoczyć, jest identyfikowanie rozmówców (często w trakcie rozmowy jest więcej niż trzy osoby). Znana jako „separacja głośników”, jest uważana za jeszcze trudniejszą niż rozpoznawanie mowy. Chorus stworzył model głębokiego uczenia się, który zasadniczo tworzy „głosowy odcisk palca” – oparty na grupowaniu – dla każdego mówcy. Tak więc po kilku latach prac badawczo-rozwojowych firma była w stanie opracować system, który mógłby analizować duże ilości rozmów. Jako dowód tego, spójrz na jednego z klientów Chorus, Housecall Pro, który jest startupem sprzedającym technologie mobilne do zarządzania usługami terenowymi. Przed wdrożeniem oprogramowania firma często tworzyła spersonalizowane oferty sprzedaży dla każdego potencjalnego klienta. Ale niestety był nieskalowalny i miał mieszane wyniki. Ale dzięki Chorusowi firma była w stanie stworzyć podejście, które nie miało zbyt wielu różnic. Oprogramowanie umożliwiło mierzenie każdego słowa i wpływu na konwersje sprzedażowe. Chorus mierzył również, czy przedstawiciel handlowy był zgodny ze scenariuszem, czy nie. Wynik? Firma była w stanie zwiększyć wskaźnik wygranych organizacji sprzedaży o 10%.

Przykład użycia: walka z depresją

Według danych Światowej Organizacji Zdrowia na całym świecie około 300 milionów ludzi cierpi na depresję.⁹ Około 15% dorosłych doświadcza pewnego rodzaju depresji w ciągu swojego życia. Może to nie zostać zdiagnozowane z powodu braku opieki zdrowotnej, co może oznaczać, że sytuacja danej osoby może się znacznie pogorszyć. Niestety depresja może prowadzić do innych problemów. Ale NLP może być w stanie poprawić sytuację. Niedawne badanie ze Stanford wykorzystywało model uczenia maszynowego, który przetwarzał mimikę 3D i język mówiony. W rezultacie system był w stanie zdiagnozować depresję ze średnim poziomem błędów 3,67 przy użyciu skali Kwestionariusza Zdrowia Pacjenta (PHQ). Dokładność była jeszcze wyższa w przypadku bardziej zastrzonych form depresji. W badaniu naukowcy zauważyli: „Ta technologia może zostać wdrożona w telefonach komórkowych na całym świecie i ułatwić tani powszechny dostęp do opieki psychiatrycznej.

Zastosowanie: tworzenie treści

W 2015 r. kilku weteranów technologii, takich jak Elon Musk, Peter Thiel, Reid Hoffman i Sam Altman, uruchomiło OpenAI przy wsparciu finansowym o wartości 1 miliarda dolarów. Celem zorganizowanej jako organizacja non-profit było stworzenie organizacji, której celem było „rozwijanie inteligencji cyfrowej w sposób, który najprawdopodobniej przyniesie korzyści całej ludzkości, nieograniczony potrzebą generowania zwrotu finansowego”. Jednym z obszarów badań był NLP. W tym celu firma uruchomiła w 2019 r. model o nazwie GPT-2, który opierał się na zbiorze danych obejmującym około ośmiu milionów stron internetowych. Skupiono się na stworzeniu systemu, który mógłby przewidzieć następane słowo na podstawie grupy tekstu.

Aby to zilustrować, OpenAI dostarczył eksperyment z następującym tekstem jako danymi wejściowymi: „W szokującym odkryciu naukowiec odkrył stado jednorożców żyjących w odległej, wcześniej niezbadanej dolinie w Andach. Jeszcze bardziej zaskakujący dla badaczy był fakt, że jednorożce mówiły doskonale po angielsku”. Na tej podstawie algorytmy stworzyły przekonującą historię o długości 377 słów! To prawda, że naukowcy przyznali, że opowiadanie historii było lepsze w przypadku tematów, które bardziej odnosiły się do podstawowych danych, takich jak Władca Pierścieni, a nawet Brexit. Nie powinno dziwić, że GPT-2 wykazał słabą wydajność w domenach technicznych. Ale model był w stanie

uzyskać wysokie wyniki w kilku dobrze znanych ocenach czytania ze zrozumieniem. Mimo że typowy człowiek uzyskałby 90%+ w tych testach, wydajność GPT-2 jest nadal imponująca. Należy zauważyć, że model wykorzystywał innowację sieci neuronowej Google, zwaną Transformer, oraz nienadzorowane uczenie się. Zgodnie z misją OpenAI organizacja zdecydowała się nie wydawać pełnego modelu. Obawiano się, że może to prowadzić do negatywnych konsekwencji, takich jak fałszywe wiadomości, sfałszowane recenzje Amazon.com, spam i oszustwa phishingowe.

Przypadek użycia: język ciała

Samo skupienie się na samym języku może być ograniczające. Język ciała to również coś, co powinno być uwzględnione w wyrafinowanym modelu sztucznej inteligencji. To jest coś, o czym Rana el Kaliouby myślała od jakiegoś czasu. Dorastając w Egipcie, uzyskała tytuł magistra nauk ścisłych na Uniwersytecie Amerykańskim w Kairze, a następnie uzyskała tytuł doktora informatyki w Newnham College na Uniwersytecie Cambridge. Ale było coś, co było dla niej bardzo przekonujące: jak komputery mogą wykrywać ludzkie emocje? Jednak w jej kręgach akademickich zainteresowanie było niewielkie. W społeczności informatyków panował zgodny pogląd, że ten temat naprawdę nie jest przydatny. Ale Rana nie zniechęciła się i połączyła siły ze znaną profesorką Rosalind Picard, aby stworzyć innowacyjne modele uczenia maszynowego (napisała kluczową książkę zatytułowaną *Affective Computing*, w której przyglądała się emocjom i maszynom). Musiało jednak być również wykorzystanie innych dziedzin, takich jak neuronauka i psychologia. Dużą częścią tego było wykorzystanie pionierskiej pracy Paula Ekmana, który przeprowadził rozległe badania nad ludzkimi emocjami opartymi na mięśniach twarzy. Odkrył, że istnieje sześć uniwersalnych ludzkich emocji (gniew, wulgaryzm, przerażenie, radość, samotność i szok), które można zakodować za pomocą 46 ruchów zwanych jednostkami działania – wszystkie stają się częścią systemu kodowania akcji twarzy (FACS). Podczas pracy w MIT Media Lab Rana opracowała „emocjonalny aparat słuchowy”, który był urządzeniem do noszenia, które umożliwiało osobom z autyzmem lepszą interakcję w środowiskach społecznych. System wykrywałby emocje ludzi i zapewniał odpowiednie sposoby reagowania. Był przełomowy, ponieważ *New York Times* nazwał go jedną z najważniejszych innowacji w 2006 roku. Ale system Rany przyciągnął również uwagę *Madison Avenue*. Mówiąc najprościej, technologia może być skutecznym narzędziem do oceny nastroju widzów w związku z reklamą telewizyjną. Kilka lat później Rana wypuściła *Affectiva*. Firma szybko się rozwinęła i przyciągnęła znaczne kwoty kapitału wysokiego ryzyka (w sumie zebrała 54,2 mln USD). Rana, kiedyś ignorowana, teraz stała się jednym z liderów trendu zwanego „sztuczną inteligencją śledzącą emocje”. Flagowym produktem *Affectiva* jest *Affdex*, czyli oparta na chmurze platforma do testowania odbiorców wideo. Korzysta z niego około jedna czwarta osób z listy *Fortune Global 500*. Ale firma opracowała inny produkt o nazwie *Affectiva Automotive AI*, który jest systemem wykrywania w kabinie pojazdu. Niektóre z możliwości obejmują:

- Monitorowanie zmęczenia kierowcy lub rozproszenia uwagi, które wywoła alarm (powiedzmy wibracje pasa bezpieczeństwa).
- Zapewnienie przełączenia do półautonomicznego systemu, jeśli kierowca się nie obudzi lub jest zły. Istnieje nawet możliwość zapewnienia alternatywnych tras, aby zmniejszyć ryzyko wściekłości na drogach!
- Personalizacja treści – powiedzmy muzyki – na podstawie emocji pasażera.

W przypadku wszystkich tych ofert istnieją zaawansowane systemy uczenia głębokiego, które przetwarzają ogromne ilości funkcji bazy danych zawierającej ponad 7,5 miliona twarzy. Modele te uwzględniają również wpływy kulturowe i różnice demograficzne – wszystko to odbywa się w czasie rzeczywistym.

Handel głosowy

Technologie oparte na NLP, takie jak wirtualni asystenci, chatboty i inteligentne głośniki, mogą mieć potężne modele biznesowe, a nawet mogą zakłócić rynki, takie jak e-commerce i marketing. Widzieliśmy już wczesną wersję tego z franczyzą WeChat firmy Tencent. Firma, która powstała w okresie rozkwitu internetowego boomu pod koniec lat 90., zaczęła od prostego komunikatora na komputery PC o nazwie OICQ. Ale to właśnie wprowadzenie WeChat zmieniło zasady gry, która od tego czasu stała się największą chińską platformą mediów społecznościowych z ponad miliardem aktywnych użytkowników miesięcznie. Ale ta aplikacja jest czymś więcej niż tylko do wymiany wiadomości i publikowania treści. WeChat szybko przekształcił się w uniwersalnego wirtualnego asystenta, w którym możesz łatwo przywołać usługę wspólnego przejazdu, dokonać płatności w lokalnym sklepie, zarezerwować lot lub zagrać w grę. Na przykład aplikacja odpowiada za blisko 35% całego czasu użytkowania na smartfonach w Chinach w ujęciu miesięcznym. WeChat jest również głównym powodem, dla którego kraj staje się coraz bardziej społeczeństwem bezgotówkowym. Wszystko to wskazuje na siłę powstającej kategorii zwanej handlem głosowym (lub v-commerce), w której można dokonywać zakupów za pośrednictwem czatu lub głosu. Jest to tak krytyczny trend, że Mark Zuckerberg z Facebooka napisał na blogu na początku 2019 roku, w którym powiedział, że firma będzie bardziej przypominać... WeChat. Według badań firmy Juniper prognozuje się, że do 2023 r. rynek handlu głosowego osiągnie 80 miliardów dolarów. Ale jeśli chodzi o zwycięzców na tym rynku, wydaje się, że to dobry zakład, że będą to firmy, które mają dużą bazę instalacji inteligentne urządzenia, takie jak Amazon, Apple i Google. Ale nadal będzie miejsce dla dostawców technologii NLP nowej generacji. W porządku, jak te systemy AI mogą wpłynąć na branżę marketingową? Cóż, żeby zobaczyć, jak, w Harvard Business Review ukazał się artykuł zatytułowany „Marketing w epoce Alexy” autorstwa Niraja Dawara i Neila Bendle'a. Autorzy zauważają w nim, że „asystenci AI zmieniają sposób, w jaki firmy łączą się ze swoimi klientami. Staną się głównym kanałem, przez który ludzie uzyskują informacje, towary i usługi, a marketing zamieni się w ciasto, które przyciąga ich uwagę”. Tak więc wzrost liczby chatbotów, asystentów cyfrowych i inteligentnych głośników może być znacznie większy niż początkowa internetowa rewolucja e-commerce. Technologie te przynoszą klientom znaczne korzyści, takie jak wygoda. Łatwo powiedzieć urządzeniu, żeby coś kupić, a maszyna dowie się również o Twoich nawykach. Więc następnym razem, gdy powiesz, że chcesz się napić, komputer będzie wiedział, o czym mówisz. Ale może to prowadzić do scenariusza, w którym zwycięzca bierze wszystko. Ostatecznie wydaje się, że konsumenci będą używać tylko jednego inteligentnego urządzenia do robienia zakupów. Ponadto marki, które chcą sprzedawać swoje towary, będą musiały dogłębnie zrozumieć, czego naprawdę chcą klienci, aby stać się preferowanym sprzedawcą w silniku rekomendacji.

Wirtualni Asystenci

W 2003 r., gdy Stany Zjednoczone były uwikłane w wojny na Bliskim Wschodzie, Departament Obrony chciał zainwestować w technologie nowej generacji na polu bitwy. Jedną z kluczowych inicjatyw było zbudowanie wyrafinowanego wirtualnego asystenta, który potrafiłby rozpoznawać instrukcje mówione. Departament Obrony przeznaczył na ten cel 150 milionów dolarów i zlecił opracowanie aplikacji laboratorium SRI (Stanford Research Institute) z siedzibą w Dolinie Krzemowej. Mimo że laboratorium było organizacją non-profit, nadal wolno mu było licencjonować swoje technologie (takie jak drukarka atramentowa) startupom. I tak właśnie stało się z wirtualnym asystentem. Niektórzy członkowie SRI - Dag Kittlaus, Tom Gruber i Adam Cheyer - nazwali ją Siri i założyli własną firmę, aby wykorzystać okazję. Założyli operację w 2007 roku, kiedy to uruchomiono iPhone'a firmy Apple.

Ale musiało być dużo więcej prac badawczo-rozwojowych, aby produkt mógł być użyteczny dla konsumentów. Założyciele musieli opracować system do obsługi danych w czasie rzeczywistym, zbudować wyszukiwarkę informacji geograficznych oraz zbudować zabezpieczenia kart kredytowych i

danych osobowych. Ale to NLP było najtrudniejszym wyzwaniem. W wywiadzie Cheyer zauważył: Najtrudniejszym technicznym wyzwaniem z Siri było radzenie sobie z ogromną ilością niejednoznaczności obecnych w ludzkim języku. Rozważ wyrażenie „zarezerwuj 4-gwiazdkową restaurację w Bostonie” – wydaje się bardzo proste do zrozumienia. Nasz prototypowy system poradziłby sobie z tym z łatwością. Kiedy jednak wprowadziliśmy do systemu dziesiątki milionów nazw firm i setki tysięcy miast jako słownictwo (prawie każde słowo w języku angielskim to nazwa firmy), liczba tłumaczeń kandydatów poszybowała w górę. Ale zespołowi udało się rozwiązać problemy i przekształcić Siri w potężny system, który został uruchomiony w App Store firmy Apple w lutym 2010 roku. „To najbardziej wyrafinowane rozpoznawanie głosu, jakie pojawiło się na smartfonie” — wynika z recenzji opublikowanej w Wired. com. Steve Jobs zwrócił na to uwagę i zadzwonił do założycieli. W ciągu kilku dni mieli się spotkać, a dyskusje szybko doprowadziły do przejęcia, które miało miejsce pod koniec kwietnia za ponad 200 milionów dolarów. Jednak Jobs uważał, że Siri wymaga ulepszeń. Z tego powodu w 2011 roku nastąpiło ponowne wydanie. To faktycznie wydarzyło się dzień przed śmiercią Jobsa. Szybko do przodu, Siri ma największy udział w rynku wirtualnych asystentów, z 48,6%. Asystent Google ma 28,7%, a Alexa Amazon.com ma 13,2%. Według raportu „Voice Assistant Consumer Adoption Report” około 146,6 mln osób w Stanach Zjednoczonych wypróbowało wirtualnych asystentów na swoich smartfonach, a ponad 50 mln z inteligentnymi głośnikami. Ale to obejmuje tylko część historii. Technologia głosowa pojawia się również w urządzeniach do noszenia, słuchawkach i urządzeniach. Oto kilka innych interesujących ustaleń:

- Używanie głosu do wyszukiwania produktów, które zostały przelicytowane w wynikach wyszukiwania różnych opcji rozrywki.
- Jeśli chodzi o produktywność, najczęstsze przypadki użycia głosu obejmują wykonywanie połączeń, wysyłanie e-maili i ustawianie alarmów.
- Najczęstsze użycie głosu w smartfonach ma miejsce, gdy osoba prowadzi samochód.
- Jeśli chodzi o skargi na asystentów głosowych na smartfonach, największym odsetkiem była niespójność w zrozumieniu prośb. Ponownie wskazuje to na ciągłe wyzwania NLP.

Potencjał wzrostu wirtualnych asystentów pozostaje jasny, a kategoria ta prawdopodobnie będzie kluczowa dla branży AI. Firma Juniper Research prognozuje, że do 2023 r. liczba wirtualnych asystentów używanych na całym świecie wzrośnie ponad trzykrotnie, do 2,5 miliarda.²⁴ Oczekuje się, że najszybszą kategorią będą telewizory inteligentne. Tak, myślę, że będziemy prowadzić rozmowy z tymi urządzeniami!

Chatboty

Często dochodzi do nieporozumień między wirtualnymi asystentami a chatbotami. Pamiętaj, że te dwa elementy w dużym stopniu się pokrywają. Obaj używają NLP do interpretacji języka i wykonywania zadań. Ale nadal istnieją krytyczne rozróżnienia. W większości chatboty są przeznaczone przede wszystkim dla firm, takich jak obsługa klienta lub funkcje sprzedażowe. Z drugiej strony, wirtualni asystenci są nakierowani zasadniczo na każdego, aby pomóc w ich codziennych czynnościach. Jak widzieliśmy w Części 1, początki chatbotów sięgają lat 60-tych wraz z rozwojem ELIZA. Ale dopiero w ostatniej dekadzie technologia ta stała się użyteczna na dużą skalę. Oto próbka interesujących chatbotów:

- Ushur: Jest to zintegrowane z systemami korporacyjnymi dla firm ubezpieczeniowych, co pozwala na automatyzację przetwarzania roszczeń/rachunków oraz umożliwia sprzedaż. Oprogramowanie wykazało średnio 30% redukcję liczby telefonów do centrum serwisowego i 90% wskaźnik odpowiedzi

klientów. Firma zbudowała własną, najnowocześniejszą lingwistykę silnik o nazwie LISA (skrót od Language Intelligence Services Architecture). LISA obejmuje NLP, NLU, analizę sentymentu, wykrywanie sarkazmu, wykrywanie tematów, ekstrakcję danych i tłumaczenia językowe. Technologia obsługuje obecnie 60 języków, dzięki czemu jest użyteczną platformą dla organizacji globalnych.

- Mya: To chatbot, który może prowadzić rozmowy w procesie rekrutacji. Podobnie jak Ushur, jest to również oparte na domowej technologii NLP. Niektóre z powodów to lepsza komunikacja, ale także zajmowanie się konkretnymi tematami dotyczącymi zatrudniania. Mya znacznie skraca czas rozmowy kwalifikacyjnej i zatrudniania, eliminując główne wąskie gardła.

- Jane.ai: jest to platforma, która wydobywa dane z aplikacji i baz danych organizacji — powiedzmy Salesforce. com, Office, Slack i Gmail — aby znacznie ułatwić uzyskiwanie spersonalizowanych odpowiedzi. Zwróć uwagę, że około 35% czasu pracownika poświęca się na szukanie informacji! Na przykład przypadkiem użycia Jane.ai jest USA Mortgage. Firma wykorzystyła technologię zintegrowaną ze Slackiem, aby pomóc brokerom w wyszukiwaniu informacji do przetwarzania kredytów hipotecznych. W rezultacie firma USA Mortgage zaoszczędziła około 1000 godzin pracy ludzkiej miesięcznie.

Mimo to chatboty nadal mają mieszane wyniki. Na przykład, tylko jednym z problemów jest to, że trudno jest programować systemy dla wyspecjalizowanych dziedzin. Spójrz na badanie przeprowadzone przez UserTesting, które zostało oparte na odpowiedziach 500 konsumentów chatbotów dla służby zdrowia. Oto niektóre z głównych wniosków: w przypadku chatbotów pozostaje wiele obaw, zwłaszcza podczas przetwarzania danych osobowych, a technologia ma problemy ze zrozumieniem złożonych tematów. Dlatego przed wdrożeniem chatbota należy wziąć pod uwagę kilka czynników:

- Ustal oczekiwania: Nie przesadzaj z możliwościami chatbotów. To tylko ustawi twoją organizację na rozczarowanie. Na przykład nie powinieneś udawać, że chatbot to człowiek. To pewny sposób na tworzenie złych doświadczeń. W rezultacie możesz zacząć rozmowę z chatbotem od słów „Cześć, jestem chatbotem, który pomoże Ci w...”

- Automatyzacja: W niektórych przypadkach chatbot może obsłużyć cały proces z klientem. Ale nadal powinieneś mieć ludzi w pętli. „Celem chatbotów nie jest całkowite zastąpienie ludzi, ale bycie pierwszą linią obrony, że tak powiem”, powiedział Antonio Cangiano, ewangelista AI w IBM. „Może to oznaczać nie tylko oszczędność pieniędzy firm, ale także uwolnienie ludzkich agentów, którzy będą mogli spędzać więcej czasu na złożonych zapytaniach, które są do nich eskalowane”.

- Tarcie: W miarę możliwości staraj się znaleźć sposób, aby chatbot rozwiązywał problemy tak szybko, jak to możliwe. I niekoniecznie musi to być rozmowa. Zamiast tego udostępnienie prostego formularza do wypełnienia może być lepszą alternatywą, powiedzmy, aby zaplanować prezentację.

- Powtarzalne procesy: są one często idealne dla chatbotów. Przykłady obejmują uwierzytelnianie, status zamówienia, planowanie i proste żądania zmian.

- Centralizacja: upewnij się, że integrujesz dane z chatbotami. Pozwoli to na bardziej płynne wrażenia. Bez wątplenia klienci szybko denerwują się, gdy muszą powtarzać informacje.

- Spersonalizuj doświadczenie: nie jest to łatwe, ale może przynieść duże korzyści. Jonathan Taylor, który jest dyrektorem ds. technologii w Zoovu, ma taki przykład: „Zakup obiektywu aparatu będzie inny dla każdego kupującego. Istnieje wiele odmian soczewek, które być może rozumie słabo poinformowany kupujący, ale przeciętny konsument może nie być tak dobrze poinformowany. Zapewnienie pomocniczego chatbota, który poprowadzi klienta do właściwego obiektywu, może

pomóc zapewnić ten sam poziom obsługi klienta, co pracownik w sklepie. Pomocniczy chatbot może zadawać właściwe pytania, rozumiejąc cel klienta, jakim jest dostarczenie spersonalizowanej rekomendacji produktu, w tym „jaki rodzaj aparatu już masz”, „dlaczego kupujesz nowy aparat” i „czego przede wszystkim próbujesz uchwycić na swoich fotografiach?”

- Analiza danych: Bardzo ważne jest monitorowanie opinii za pomocą chatbota. Jaka jest satysfakcja? Jaki jest wskaźnik dokładności?
- Projektowanie konwersacji i doświadczenie użytkownika (UX): to coś innego niż tworzenie strony internetowej czy nawet aplikacji mobilnej. Z chatbotem musisz pomyśleć o osobowości użytkownika, płci, a nawet kontekście kulturowym. Co więcej, musisz wziąć pod uwagę „głos” swojej firmy. „Zamiast tworzyć makiety interfejsu wizualnego, pomyśl o pisaniu skryptów i odtwarzaniu ich przed jego zbudowaniem” – powiedziała Gillian McCann, szefowa działu inżynierii chmury i sztucznej inteligencji w Workgrid Software.

Nawet w przypadku problemów z chatbotami technologia wciąż się poprawia. Co ważniejsze, systemy te prawdopodobnie staną się coraz ważniejszą częścią branży AI. Według IDC w 2019 r. na chatboty zostanie wydanych około 4,5 miliarda dolarów, co w porównaniu z 35,8 miliardami dolarów szacowanymi na systemy sztucznej inteligencji. Coś jeszcze: badanie przeprowadzone przez Juniper Research wskazuje, że oszczędności dzięki chatbotom mogą być znaczne. Firma przewiduje, że do 2023 r. osiągną 7,3 mld USD, w porównaniu z zaledwie 209 mln USD w 2019 r.³³.

Przyszłość NLP

W 1947 roku Boris Katz urodził się w Mołdawii, która była częścią Związku Radzieckiego. Następnie ukończył Moskiewski Uniwersytet Państwowy, gdzie poznał komputery, a następnie wyjechał do Stanów Zjednoczonych (z pomocą senatora Edwarda Kennedy'ego). Nie marnował czasu przy tej okazji. Oprócz napisania ponad 80 publikacji technicznych i uzyskania dwóch patentów w USA, stworzył system START, który pozwalał na zaawansowane możliwości NLP. Była to właściwie podstawa pierwszej witryny internetowej z pytaniami i odpowiedziami w 1993 roku. Tak, była to zapowiedź przełomowych firm, takich jak Yahoo! i Google. Innowacje Borisa miały również kluczowe znaczenie dla Watsona IBM, który jest obecnie podstawą działań firmy w zakresie sztucznej inteligencji. Ten komputer w 2011 roku zaszokował świat, pokonując dwóch wszechczasów mistrzów popularnego teleturnieju Jeopardy! Pomimo całego postępu w NLP, Boris nie jest zadowolony. Wierzy, że wciąż jesteśmy na wczesnym etapie i trzeba zrobić dużo więcej, aby uzyskać prawdziwą wartość. W wywiadzie dla MIT Technology Review powiedział: „Ale z drugiej strony te programy [takie jak Siri i Alexa] są tak niewiarygodnie głupie. Jest więc poczucie dumy i niemal zakłopotania. Wprowadzasz coś, co ludzie uważają za inteligentne, ale nawet nie jest blisko”. Nie oznacza to, że jest pesymistą. Jednak nadal uważa, że należy przemyśleć NLP, jeśli ma dojść do „prawdziwej inteligencji”. Uważa, że w tym celu badacze muszą wykroczyć poza czystą informatykę do szerokich obszarów, takich jak neuronauka, kognitywistyka i psychologia. Uważa również, że systemy NLP muszą znacznie lepiej rozumieć działania w prawdziwym świecie.

Wniosek

Dla wielu osób pierwsza interakcja z NLP odbywa się z wirtualnymi asystentami. Nawet jeśli technologia jest daleka od doskonałości, nadal jest bardzo przydatna - zwłaszcza do odpowiadania na pytania lub uzyskiwania informacji, powiedzmy o pobliskiej restauracji. Ale NLP ma również duży wpływ na świat biznesu. W nadchodzących latach technologia będzie stawać się coraz ważniejsza dla handlu elektronicznego i obsługi klienta – zapewniając znaczne oszczędności kosztów i pozwalając pracownikom skupić się na działaniach o większej wartości dodanej. To prawda, że ze względu na

złożoność języka przed nami jeszcze długa droga. Ale postęp jest nadal szybki, zwłaszcza z pomocą podejść AI nowej generacji, takich jak głębokie uczenie.

Kluczowe dania na wynos

- Przetwarzanie języka naturalnego (NLP) to wykorzystanie sztucznej inteligencji w celu umożliwienia komputerom zrozumienia ludzi.
- Chatbot to system AI, który komunikuje się z ludźmi, powiedzmy za pomocą głosu lub czatu online.
- Chociaż w NLP poczyniono wielkie postępy, jest jeszcze wiele do zrobienia. Tylko niektóre z wyzwań obejmują niejednoznaczność języka, sygnały niewerbalne, różne dialekty i akcenty oraz zmiany w języku.
- Dwa główne kroki z NLP obejmują czyszczenie/wstępne przetwarzanie tekstu i wykorzystanie sztucznej inteligencji do zrozumienia i wygenerowania języka.
- Tokenizacja polega na analizowaniu tekstu i dzieleniu go na różne części.
- Dzięki normalizacji tekst jest przekształcany w formę, która ułatwia analizę, na przykład przez usunięcie znaków interpunkcyjnych lub skrótów.
- Stemming opisuje proces redukowania słowa do jego rdzenia (lub lematu), na przykład poprzez usuwanie afiksów i przyrostków.
- Podobnie do tematu, lematyzacja polega na znalezieniu podobnych słów źródłowych.
- Aby NLP mógł zrozumieć język, istnieją różne podejścia, takie jak tagowanie części mowy (umieszczanie tekstu w formie gramatycznej), chunking (przetwarzanie tekstu we frazach) i modelowanie tematu (znajdowanie ukrytych wzorców i klastrów).
- Fonem to najbardziej podstawowa jednostka dźwiękowa w języku.